STANFORD UNIVERSITY

CENTER FOR SYSTEMS RESEARCH

STATE ESTIMATION WITH SMALL NON-LINEARITIES

by

BJÖRN CONRAD

MARCH 1971

Fill also

Guidance and Control Laboratory

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

AFAL-TR-71-222

STATE ESTIMATION WITH SMALL NON-LINEARITIES

By

Björn Conrad

March 1971

Approved for public release;
distribution unlimited.

SUDAAR No. 421

This work was performed in part for NASA
under Contract NGR-05-020-019 and the Air Force
under Contracts F33615-C-67-1245 and F33615-C-70-1637

Department of Aeronautics and Astronautics
Stanford University
Stanford, California

FOREWORD

The work reported here was performed under Contracts F33615-67-C-1245 and F33615-70-C-1637, Project 5102, Task 510215 at the Center for Systems Research, Guidance and Control Laboratory, Stanford, California.

This report has been issued as Department of Aeronautics and Astronautics document SUDAAR NO. 421.

This program is monitored by Mr. H. P. Kohlmann, Air Force Avionics Laboratory (AFAL/NVE).

Publication of this report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

This technical report has been reviewed and is approved for publication.

WILLIAM J. DELANEY
Lt Colonel, USAF
Chief, Navigation &
Guidance Division

ABSTRACT

A variety of techniques are available for estimating the states of non-linear dynamic systems from noisy data. These procedures are generally equivalent when applied to linear systems. This dissertation investigates the difference between several of these procedures in the presence of small dynamic and observational non-linearities.

Four discrete estimation algorithms are analyzed. The first is a strictly least square estimator, while the other three are recursive algorithms similar to the Kalman filter used for estimating the states of linear systems. The product of this research is a group of analytic expressions for the mean and covariance of the error in each of these estimators so that they may be compared without lengthy Monte-Carlo simulations.

The covariance expressions show that, to first order, all the estimators have the same covariance. Expressions for the means, however, show that each estimator has a different bias. Several examples are carried out demonstrating that the relative magnitudes of the bias errors in the various estimators can be a strong function of such parameters as initial covariances and number of data points being considered. In fact, under some circumstances it appears that more complicated (seemingly superior) algorithms can have larger biases than smaller ones.

CONTENTS

PRECEDING PAGE BLANK NOT FILMED

CONTENTS (Cont)

TABLES

ILLUSTRATIONS

ACKNOWLEDGMENTS

The author wishes to thank Professors Benjamin O. Lange and John V. Breakwell for their support in this research task, and Professor Thomas Kailath for his criticism of the manuscript. Special thanks are also due my wife Joyce whose encouragement and assistance helped make this work possible.

Partial support for this work has come from NASA under Contract NGR-05-020-019 and the Air Force under Contracts F33615-C-67-1245 and F33615-C-70-1637.

1. INTRODUCTION

1.1 Statement of the Problem

In the last decade, much attention has been focused on the problem of extracting information about a system from noisy measurement data. Work in the area of state estimation for linear dynamic systems has yielded especially useful results, producing a good understanding of the overall problem as well as effective general purpose procedures for processing noisy data from such systems. Comparable results have not been obtained for broad classes of non-linear dynamic systems, but experience has shown good state estimates of many such systems can be realized by using algorithms based on various extensions of the linear theory. A goal of this dissertation is to find an analytic basis for a comparison between some of the more popular algorithms as applied to a class of multi-dimensional discrete dynamic systems which can be characterized as "slightly non-linear".

In the scalar case these "slightly non-linear" systems are described by a state x_k which satisfies the difference equation

$$x_{k+1} = \varphi_k x_k + \epsilon_k x_k^2 + q_k \qquad k = 0,1,2, \ldots, N$$

and noisy measurements y_k assumed to obey the observation equation

$$y_k = h_k x_k + e_k x_k^2 + \nu_k$$

The parameters φ_k, h_k, ϵ_k and e_k are assumed known, and the problem is such that terms involving the second two parameters can be assumed small. The sequences q_k and ν_k are random variables constituting "noise" and x_0 is an unknown random variable. Precise definitions of all these quantities is deferred to Chapter 2 which contains a description of the multi-dimensional slightly non-linear system.

Given a sequence of data y_i, $i = 1,2, \ldots, N$, from such a system, the following algorithms might be used to estimate x_N:

1. The classical least square fitting procedure (LS algorithm)

2. An extended Kalman Filter obtained by linearizing about a predicted state based on the current best estimate (EK algorithm)

3. A procedure similar to Kalman's but based on making approximate step by step minimum variance estimates up to first order (AM algorithm)

4. An iterative algorithm similar to (2) above but linearizing about a current best estimate, thus requiring a simple iteration (IT algorithm).

These algorithms are also defined explicitly in Chapter 2. The third one presents a slight difficulty, since it has not been derived for discrete systems. Assuming any of these algorithms will give a reasonable estimate of x_N for a particular system, an analyst may choose to compare them in terms of the bias or error covariance histories they imply. This dissertation seeks to find approximate analytic expressions for these histories.

The purpose of such an analysis is twofold. First, most systems are at least slightly non-linear. In the case of very precise data fitting requirements it would be desirable to have expressions which predict the possible effects of these non-linearities on a particular estimation scheme. For example, Stanford University is considering measuring a small relativistic phenomenon via an orbiting gyroscope and it is important that there are no errors in the estimation algorithm which could dwarf the parameters being determined.

Second, it is hoped that an understanding of the slightly non-linear estimation problem can lend some intuition to the limitations of these common estimation schemes for larger non-linearities which often occur when measurements are complicated (often geometric) functions of the states and the system random variables contain large uncertainties. This situation frequently arises when only a few observations from a system are available.

1.2 Previous Results

In the last ten years more analytic effort has been devoted toward finding optimal general purpose filters for continuous non-linear filters than on analyzing currently available algorithms.* Much of this work is typified by Bass et.al. (Ref. 4). Their emphasis on continuous (as opposed to discrete) systems is motivated by the desire to utilize continuity properties to obtain diffusion equations describing the relevant conditional probabilities on which any optimal state estimate must depend. Major drawbacks to much of this work have been:

a. the necessity of dealing with complicated mathematical concepts associated with continuous non-linear functions of stochastic processes

b. the production of solutions in terms of hard to solve partial differential equations

c. the difficulty in interpreting the validity of approximate solutions to the partial differential equations in terms of constraints on real problems and

d. the lack of explicit information on whether filters so derived will yield advantages over better known, and often simpler, procedures.

In 1966, Jazwinski (Ref. 9) unsuccessfully attempted** to derive a discrete analogue to the non-linear filters discussed above. Two years

---

* Much of this work appears to be motivated by attempts to find filtering solutions which are both "optimal" and straightforwardly practical in the same sense as those of Kalman's, e.g. (Ref. 10) for linear systems. For practical purposes, however, the concept of optimality is often of secondary importance since the statistics, and indeed the whole model, are seldom known precisely. In fact, the general form of the Kalman filter has been used to solve many problems without satisfying any of the requirements necessary for it to be optimal. The chief benefit of Kalman's work appears to be its emphasis on the relationship between the structure of the linear systems producing data and their corresponding filters.

** The writer has been able to isolate a fundamental error in Jazwinski's approach. In fact, R. Curran, a research assistant at Stanford University, found Jazwinski's algorithm to be unstable for some simple problems.

later, Athans et. al. (Ref. 2) presented a partial discrete analogue to the approximate non-linear filters of Bass et. al. The word "partial" is used because they did not consider state noise and the dynamics were assumed continuous while observations were assumed discrete. A drawback of their derivation appears to be that terms which turned out to be empirically negligible were not recognized as such. Athans used Monte Carlo techniques to evaluate his filter for a particular problem.

Breakwell, in Ref. 5, undertook the task of analyzing several popular estimation schemes as applied to slightly non-linear parameter estimation problems. In particular, he derived approximate relations for the mean and covariance of the least-square filter, the extended Kalman filter and an iterated version of the latter. These relations produced the surprising result that, although the least-square algorithm was always best, the iterated filter could have larger bias errors than the simpler extended Kalman filter. Furthermore, he found that all the estimators (to first order) had the same covariance.

1.3 New Results

The following objectives were achieved in this research effort. First, a simple development of a second order discrete analogue of the continuous filters of Bass, et. al. has been found. This filter is slightly more general than that of Athans, and the terms which they found to be negligible empirically do not occur. This filter is developed in Section 2.4 and analyzed in Section 3.3.

Second, approximate equations have been found for comparing the difference between the bias and covariance of four estimators: the least-square (LS); the extended Kalman (EK); the second order approximate minimum variance (AM); and the iterated extended Kalman (IT). An explicit mathematical development of each of these algorithms is contained in Chapter 2, along with a precise definition of the slightly non-linear systems for which they were analyzed. The actual analysis of these algorithms in terms of bias and covariance expressions is carried out in Chapter 3.

Although each estimator has a different bias history (to first order), their covariances all turn out to be the same. Furthermore, the only difference between the slightly non-linear covariance, and one computed ignoring all non-linearities, is a term linearly related to the *a priori* mean of the initial state x. If the model is chosen so that this mean is zero, the non-linear covariance is the same as the linear. These equations should circumvent the need for Monte-Carlo comparisons of these estimators for slightly non-linear problems. Furthermore, it is hoped that they will provide intuition into the effects of larger non-linearities on the state estimation problem.

Third, some examples are given in Chapter 4 which point out interesting contrasts and similarities between this work and the results obtained by Breakwell for the special case of slightly non-linear parameter estimation. These examples show how the nature of *a priori* information, as well as the value of initial covariances, can determine which estimator has the largest bias for a given number of observations. A dynamic example is also presented.

PRECEDING PAGE BLANK NOT FILMED

2. STATE ESTIMATION ALGORITHMS FOR SLIGHTLY NON-LINEAR SYSTEMS

Introduction

This chapter presents a mathematical description of the slightly non-linear systems and associated data fitting algorithms with which this dissertation is concerned. It begins in Section 2.1 by explicitly defining the slightly non-linear dynamic system whose noisy output data must be processed to obtain information about its state. It then goes on to develop four computational algorithms for processing this data to obtain the required information.

The processing schemes developed in Section 2.2 through 2.4, respectively, are:

1. the classical least square fitting procedure (LS) resulting in a two point boundary value problem (TPBVP) which must be solved by iterative techniques
2. the extended Kalman filter (EK), derived seeking an approximate sequential estimator for the least square procedure
3. the approximate minimum-variance estimator (AM) which differs from 2 above in that it tries to minimize an expected value of the estimate, rather than just a function of given data
4. the iterated (IT) filter which is a refinement of the EK filter.

It should be noted that the development of these algorithms is somewhat heuristic, and the assumptions used to derive them are not necessarily identical with those used in the analysis of the following chapter.

2.1 Discrete System Description

A fairly general non-linear discrete system can be described by the forward difference equation

$$x_{k+1} = f_k(x_k) + G_k q_k \tag{2.1}$$

with output observation

$$y_k = h_k(x_k) + \nu_k \tag{2.2}$$

where the x_k is a sequence of $n \times 1$ matrices, q_k is a sequence of $p \times 1$ matrices, and the y_k and ν_k constitute a sequence of $m \times 1$ matrices. f_k and h_k are at least three times differentiable matrix functions and the q_k and ν_k are zero mean processes with covariances:*

$$\begin{gathered} E\left\{q_k q_j^T\right\} = Q_k \delta_{kj} \; ; \qquad E\left\{\nu_k \nu_j^T\right\} = R_k \delta_{kj} \\ E\left\{\nu_k q_j\right\} = 0 \; ; \qquad E\left\{x_k \nu_j\right\} = 0 \; ; \qquad E\left\{x_k q_k\right\} = 0 \end{gathered} \tag{2.3}$$

Furthermore, the initial conditions on (2.1) are assumed to be random variables with <u>a priori</u> statistics

$$\begin{gathered} E\left\{x_0\right\} = \bar{x}_0 \\ E\left\{\left(x_0 - \bar{x}_0\right)\left(x_0 - \bar{x}_0\right)^T\right\} = P_0 \end{gathered} \tag{2.4}$$

The development in the sequel will be limited to estimating the x_k given the y_k in the cases where Taylor series expansions of (2.1) and (2.2) can be performed. When these equations are expanded in a Taylor series about some nominal operating point $\hat{x}_k$ we can define new variables in terms of differences

$$\begin{gathered} \Delta x_k \triangleq x_k - \hat{x}_k \\ \Delta y_k \triangleq y_k - \hat{y}_k \end{gathered} \tag{2.5}$$

to obtain

*E, as used here, is the expectation operator.

$$\Delta x_{k+1} = \frac{\partial f_k}{\partial x_k}(\hat{x}_k)\,\Delta x_k + \left[\frac{1}{2}\,\frac{\partial^2 f_k}{\partial^2 x_k} : \Delta x_k \Delta x_k^T\right] + G_k q_k \tag{2.6a}$$

$$+ \text{ terms with } (\Delta x)^3, \quad (\Delta x)^4 \text{ etc.}$$

and

$$\Delta y_k = \frac{\partial h}{\partial x_k}(\hat{x}_k)\,\Delta x_k + \left[\frac{1}{2}\,\frac{\partial^2 h_{xx}}{\partial x_k^2} : \Delta x_k \Delta x_k^T\right] + \nu_k \tag{2.6b}$$

$$+ \text{ terms with } \Delta x^3, \quad \Delta x^4 \text{ etc.}$$

In view of the fact that q_k and ν_k are unknown random variables which, along with the x_k can never be determined exactly from the data y_k, it is difficult to find the most general conditions which assure that a close enough $\hat{x}_k$ can ever be found to validate the use of these expansions. One way to skirt this difficulty is to make the strong assumption that the probability distributions associated with all the random variables in the problem are bounded and have small enough variances to make the expansions tractable. The necessity of investigating these assumptions for particular problems is not well emphasized in the literature dealing with continuous non-linear dynamic systems where more attention is focused on theoretical aspects of continuous stochastic processes.

For future convenience it will be assumed that expansions of the form (2.6) and (2.7) are valid so that by scaling and redefining state variables the only system which need be considered is

$$x_{k+1} = \Phi_k x_k + \left[\mathcal{E}_k : x_k x_k^T\right] + G_k q_k^{\dagger} \tag{2.7a}$$

---

† An important consequence of this notation is that terms with coefficients of order $\mathcal{E}_k \Phi_k$ must be neglected along with $\mathcal{E}_k^2$ when performing algebraic manipulations.

$$y_k = H_k x_k + \left[E_k : x_k x_k^T\right]^* + \nu_k \tag{2.7b}$$

with the statistics (2.3) and (2.4), together with the assumption that $\mathcal{E}_k$ and E_k are small so that squares and higher order products of these terms are negligible.

2.2 Classical Least Square Fitting

The preceding section defined the source of the data which must be processed. Of the four algorithms (the LS, EK, AM, and IT) to be considered for this task the least square will be presented first since it is the most popular, especially in problems with few computational constraints. In this approach, estimates of the x_k and q_k are made by determining a sequence $\hat{x}_k$ and $\hat{q}_k$ which obey the constraint relations

$$\begin{aligned} \hat{x}_{k+1} &= \Phi_k \hat{x}_k + \left[\mathcal{E}_k : \hat{x}_k \hat{x}_k^T\right] + G_k \hat{q}_k \qquad k = 0, 1, \ldots, N-1 \\ \hat{y}_k &= H_k \hat{x}_k + \left[E_k : \hat{x}_k \hat{x}_k^T\right] \end{aligned} \tag{2.8}$$

and minimize the quadratic performance index

$$\varphi = \frac{1}{2}\left(\hat{x}_0 - \bar{x}_0\right)^T P_0^{-1}\left(\hat{x}_0 - \bar{x}_0\right) + \frac{1}{2}\sum_{i=1}^{N}\left[\left(y_i - \hat{y}_i\right)^T R_i^{-1}\left(y_i - \hat{y}_i\right) + \hat{q}_{i-1}^T Q_{i-1}^{-1}\hat{q}_{i-1}\right] . \tag{2.9}$$

Since N is finite, this is essentially an algebraic maximization problem, solvable by using (2.8) to eliminate all the intermediate states $\hat{x}_i$ with i not equal to 0. However, this leads to the difficulty that all the $\hat{q}_i$ must be found explicitly. This is avoided by adjoining

---

*Notational definitions;

$[E:P]_u \triangleq \sum_{\alpha,\beta} E_{u\alpha\beta} P_{\alpha\beta}$; $[E:x]_{uv} \triangleq \sum_{\alpha} E_{uv\alpha} x_\alpha$; $[E^T:P]_u \triangleq \sum_{\alpha,\beta} E_{\alpha\beta u} P_{\alpha\beta}$

the constraints (2.8) to (2.9) via a sequence of Lagrange multipliers λ_k so that an augmented performance index is written

$$\bar{\varphi} = \varphi + \sum_{i=0}^{N-1} \lambda_i^T \left(\Phi_i \hat{x}_i + \left[\mathcal{E}_i : \hat{x}_i \hat{x}_i^T \right] + G_i \hat{q}_i - \hat{x}_{i+1} \right) \quad . \tag{2.10}$$

Differentiating (2.10) with respect to $\hat{q}_k$ and setting the result equal to zero to find an extremum gives

$$\hat{q}_k = -Q_k G_k^T \lambda_k \qquad k = 0, 1, \ldots, N-1 \quad . \tag{2.11}$$

Repeating this process for $\hat{x}_k$ gives:

$$\lambda_{k-1} = \left(\Phi_k^T + 2 \left[\mathcal{E}_k : x_k \right]^T \right) \lambda_k - \left(H_k^T + 2 \left[E_k : x_k \right]^T \right) R_k^{-1} \left(y_k - \hat{y}_k \right) \tag{2.12}$$

$$k = 1, 2, \ldots, N$$

with

$$\lambda_N = 0 \quad . \tag{2.13}$$

Differentiating (2.10) with respect to $\hat{x}_0$ gives the initial boundary condition

$$-P_0^{-1} \left(\bar{x}_0 - \hat{x}_0 \right) + \left(\Phi_0^T + 2 \left[\mathcal{E}_0 : \hat{x}_0 \right]^T \right) \lambda_0 = 0 \quad . \tag{2.14}$$

Equation (2.11) can be substituted into (2.8) to obtain the final equation:

$$\hat{x}_{k+1} = \Phi_k \hat{x}_k + \left[\mathcal{E}_k : \hat{x}_k \hat{x}_k^T \right] - G_k Q_k G_k^T \lambda_k \tag{2.15}$$

Equations (2.12) through (2.15) constitute a discrete two point boundary value problem (TPBVP) whose solution constitutes the classical least square smoothing estimate.

2.3 The Kalman Recursive Algorithms Generalized to Non-linear Systems (Extended Kalman Filter)

Although the recursive filtering equations developed by Kalman (Ref. 10) are only valid for linear systems, they can often be applied to non-linear systems which have been linearized about some reference path. The algorithm that will be considered here is obtained by considering a reference path which at any time is linearized about a current best estimation of a state. The resultant equations look much like the Kalman equations with linear transformations replaced by their non-linear equivalents.

As a first step in obtaining this algorithm we will assume that a good approximation to a portion of the least square problem represented by Eqs. (2.8) and (2.9) has been found. This solution is expressed as $\hat{x}_{\ell/\ell}$ where the second index indicates that the $\hat{x}$ so defined is part of the sequence satisfying Eqs. (2.8) minimizing the partial performance index

$$\varphi_\ell = \frac{1}{2}\left(\hat{x}_0 - \bar{x}_0\right)^T P_0^{-1}\left(\hat{x}_0 - \bar{x}_0\right) + \frac{1}{2}\sum_{i=1}^{\ell}\left[\left(y_i - \hat{y}_i\right)^T R_i^{-1}\left(y_i - \hat{y}_i\right) + \hat{q}_{i-1}^T Q_{i-1}^{-1}\hat{q}_{i-1}\right] \tag{2.16}$$

$$\ell \leq N$$

If indeed $\hat{x}_{\ell/\ell}$ is a good approximation to a minimum, we know that the derivative of (2.16) with respect to the components of $\hat{x}_{\ell/\ell}$ must be zero. Hence a Taylor series expansion of φ_ℓ about $\hat{x}_{\ell/\ell}$ can be written as

$$\varphi_\ell^* = \varphi^*\left(\hat{x}_{\ell/\ell}\right) + \frac{1}{2}\,\delta\hat{x}_{\ell/\ell+1}^T\,(P_\ell^*)^{-1}\,\delta\hat{x}_{\ell/\ell+1} + \text{H.O.T.}^{\dagger} \tag{2.17}$$

† The asterisk is used to denote quantities associated with an optimum solution.

where $(P^*_\ell)^{-1}$ is presumed known, $\delta\hat{x}_{\ell/\ell+1}$ is a small excursion from $\hat{x}_{\ell/\ell}$ defined by

$$\delta\hat{x}_{\ell/\ell+1} = \hat{x}_{\ell/\ell+1} - \hat{x}_{\ell/\ell} \tag{2.18}$$

and $\hat{x}_{\ell/\ell+1}$ represents the estimate of x_ℓ obtained by minimizing

$$\varphi_{\ell+1} = \varphi_\ell + \frac{1}{2}\left(y_{\ell+1} - \hat{y}_{\ell+1}\right)^T R^{-1}_{\ell+1}\left(y_{\ell+1} - \hat{y}_{\ell+1}\right) + \frac{1}{2}\hat{q}^T_\ell Q^{-1}_\ell \hat{q}_\ell \tag{2.19}$$

In order to obtain a simple forward solution for $\hat{x}_{\ell+1/\ell+1}$ as a function of $\hat{x}_{\ell/\ell}$ and $y_{\ell+1}$ it is assumed that the change (2.18) obtained by minimizing (2.19) rather that (2.16) is small so that the approximation (2.17) can be used in (2.19). Furthermore, it will also be assumed that $\hat{q}_\ell$ will also be small. Taken with the above assumptions, this implies that $\hat{x}_{\ell+1/\ell}$ will be close to the predicted value defined by:

$$\hat{x}_{\ell+1/\ell} = \Phi_\ell \hat{x}_{\ell/\ell} + \left[\mathcal{E}_\ell : \hat{x}_\ell \hat{x}^T_\ell\right] . \tag{2.20}$$

Hence, it follows that

$$\delta\hat{x}_{\ell+1/\ell+1} \triangleq \hat{x}_{\ell+1/\ell+1} - \hat{x}_{\ell+1/\ell} \tag{2.21}$$

should also be small.† This last assumption permits us to expand the second term on the right side of Eq. (2.19) about $\hat{x}_{\ell+1/\ell}$. Recalling Eq. (2.8), this expansion is performed by writing

$$\begin{aligned} y_{\ell+1} - \hat{y}_{\ell+1} \cong y_{\ell+1} &- \left(H_{\ell+1}\hat{x}_{\ell+1/\ell} + \left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\hat{x}^T_{\ell+1/\ell}\right]\right) \\ &- \left(H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right]\right)\delta\hat{x}_{\ell+1/\ell+1} \end{aligned} \tag{2.22}$$

† The hypothesis here is that the E, δ, and q terms are of the same order.

It might be argued that when (2.22) is substituted into (2.19), quadratic terms in $\delta\hat{x}_{\ell+1/\ell+1}$ should also be included since (2.17) contains quadratic terms. Further expansion of (2.22), however, involves only the products of quadratics with $E_{\ell+1}$, which, by hypothesis, are second order terms and hence negligible in magnitude.

In order to relate $\delta\hat{x}_{\ell+1/\ell+1}$ in (2.22) to $\delta\hat{x}_{\ell/\ell+1}$ found in (2.17), the defining relations (2.8), (2.18), (2.20) and (2.21) are combined to give a new constraint relation

$$\delta\hat{x}_{\ell+1/\ell+1} \cong \Phi_\ell \delta\hat{x}_{\ell/\ell+1} + G_\ell \hat{q}_\ell \quad . \tag{2.23}$$

Before redefining the least square problem in terms of $\delta\hat{x}$ it is convenient to define two new variables:

$$y'_{\ell+1} \triangleq y_{\ell+1} - H_{\ell+1}\hat{x}_{\ell+1/\ell} - \left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\hat{x}^T_{\ell+1/\ell}\right] = y_{\ell+1} - \hat{y}_{\ell+1/\ell} \tag{2.24}$$

and

$$H'_{\ell+1} = H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right] \tag{2.25}$$

Using these definitions, the approximations (2.22) and (2.17) are substituted into (2.19) to obtain the minimization problem that is now summarized. Minimize

$$\begin{aligned}\varphi_{\ell+1} &= \varphi^*_\ell\left(\hat{x}_{\ell/\ell}\right) + \frac{1}{2}\,\delta\hat{x}^T_{\ell/\ell+1}(P^*_\ell)^{-1}\,\delta\hat{x}_{\ell/\ell+1} \\ &\quad + \frac{1}{2}\left(y'_{\ell+1} - H'_{\ell+1}\delta\hat{x}_{\ell+1/\ell+1}\right)^T R^{-1}_{\ell+1}\left(y'_{\ell+1} - H'_{\ell+1}\delta\hat{x}_{\ell+1/\ell+1}\right) + \frac{1}{2}\,\hat{q}^T_\ell Q^{-1}_\ell \hat{q}_\ell\end{aligned} \tag{2.26}$$

over $\hat{q}_\ell$, $\delta\hat{x}_{\ell/\ell+1}$ and $\delta\hat{x}_{\ell+1/\ell+1}$ subject to

$$\delta\hat{x}_{\ell+1/\ell+1} = \Phi_\ell \delta\hat{x}_{\ell/\ell+1} + G_\ell \hat{q}_\ell \tag{2.27}$$

But these equations simply represent one step of the more general problem solved in the appendices. The solution is simply

$$\delta\hat{x}_{\ell+1/\ell+1} = P^*_{\ell+1} H'^T_{\ell+1} R^{-1}_{\ell+1} y'_{\ell+1} \tag{2.28}$$

$$(P^*_{\ell+1})^{-1} \triangleq \left(\Phi_\ell P^*_\ell \Phi^T_\ell + G_\ell Q_\ell G^T_\ell\right)^{-1} + H'^T_{\ell+1} R^{-1}_{\ell+1} H'_{\ell+1} \tag{2.29}$$

The complete recursive algorithm can now be summarized. Equation (2.28) will be combined directly with (2.21) to eliminate explicit computation of the perturbation term. Hence, given $\hat{x}_{\ell/\ell}$ and P^*_ℓ, the $\ell+1^{th}$ terms are computed from:†

$$\hat{x}_{\ell+1/\ell} = \Phi_\ell \hat{x}_{\ell/\ell} + \left[\mathcal{E}_\ell : \hat{x}_\ell \hat{x}^T_\ell\right] \tag{2.30a}$$

$$H'_{\ell+1} = H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right] \tag{2.30b}$$

$$P^*_{\ell+1/\ell} = \Phi_\ell P^*_\ell \Phi^T_\ell + G_\ell Q_\ell G^T_\ell \tag{2.30c}$$

$$(P^*_{\ell+1})^{-1} = (P^*_{\ell+1/\ell})^{-1} + H'^T_{\ell+1} R^{-1}_{\ell+1} H'_{\ell+1} \tag{2.30d}$$

† It should be noted that the state prediction equation (2.30a) for $\hat{x}_{\ell+1/\ell}$ and the predicted observation used in (2.30e) can be replaced by:

$$\hat{x}_{\ell+1/\ell} = f_\ell\left(\hat{x}_{\ell/\ell}\right)$$

and

$$\hat{y}_{\ell+1/\ell} = h_{\ell+1}\left(\hat{x}_{\ell+1/\ell}\right)$$

for the more general problem discussed in Section 2.1. Furthermore, the inverse required in Eq. (2.30d) can be eliminated by using the well known relation

$$P^*_{\ell+1} = P^*_{\ell+1/\ell} - P^*_{\ell+1/\ell} H'^T_{\ell+1} \left(H'_{\ell+1} P^*_{\ell+1/\ell} H'^T_{\ell+1} + R_{\ell+1}\right)^{-1} H'_{\ell+1} P^*_{\ell+1/\ell}$$

$$\hat{y}_{\ell+1/\ell} = H_{\ell+1}\hat{x}_{\ell+1/\ell} + \left[E_{\ell+1}:\hat{x}_{\ell+1/\ell}\hat{x}^T_{\ell+1/\ell}\right] \tag{2.30e}$$

$$\hat{x}_{\ell+1/\ell+1} = \hat{x}_{\ell+1/\ell} + P^*_{\ell+1}H'^T_{\ell+1}R^{-1}_{\ell+1}\left(y_{\ell+1} - \hat{y}_{\ell+1/\ell}\right) \tag{2.30f}$$

$$\hat{x}_0 = \bar{x}_0 \tag{2.30g}$$

This concludes the development of what is frequently referred to as the extended Kalman filter.

2.4 An Approximate Minimum Variance Filter

It is well known (Refs. 13 and 14) that the minimum variance estimates of a random variable such as x_k based on data such as y_i is the conditional mean of x_k given the y_i $(i = 1,2, \ldots, N)$. In this section an approximate recursive algorithm for computing the conditional mean is derived. At several points in the derivation, certain variables are arbitrarily assumed to have gaussian distributions. The analysis of this algorithm in Chapter 3, however, does not rely on this assumption. Thus, the algorithm has greater applicability than some of the intermediate approximations might, at first hand, indicate.

The derivation begins with the assumption that x_k has been estimated based on the data y_i, $(i = 1,2, \ldots, N)$. This estimate will be called $\hat{x}_k$ and is presumed to be the conditional mean of x_k written symbolically as

$$\hat{x}_k = E\left\{x_k/y_1, \ldots, y_k\right\} = \int_{-\infty}^{\infty} x_k p(x_k/y_1, \ldots, y_k)\, dx_k \tag{2.31}$$

In this equation p is the conditional multivariate probability density of x_k conditioned on the data y. The integral sign and differential are symbolic and should be interpreted as

$$\int_{-\infty}^{\infty} dx_k = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\left(x_k\right)_1, d\left(x_k\right)_2, d\left(x_k\right)_3, \ \ldots, \ d\left(x_k\right)_N \ ,$$

that is as an integral over all the components of x_k. Furthermore, it will also be assumed that the covariance of the error in this estimate is also known. This covariance is defined by

$$E\left\{\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right\} \triangleq E\left\{\tilde{x}_k\tilde{x}_k^T\right\} = \bar{P}_k \tag{2.32}$$

Given the above information we consider first the problem of finding the minimum variance estimate of x_{k+1} given only the data up to y_k. This estimate is denoted by $\hat{x}_{k+1/k}$ and the corresponding covariance by $P_{k+1/k}$. By definition,

$$\hat{x}_{k+1/k} = E\left\{x_{k+1}/y_1, \ \ldots, \ y_k\right\} = E\left\{\Phi_k x_k + \left[\mathcal{E}_k : x_k x_k^T\right] + G_k q_k/y_1, \ \ldots, \ y_k\right\} \tag{2.33}$$

By hypothesis, q_k is a zero mean process which is uncorrelated with data up to and including y_k. Furthermore, the expectation operator is linear, so we can write (2.33) as

$$\hat{x}_{k+1/k} = \Phi_k E\left\{x_k/y_1, \ \ldots, \ y_k\right\} + \left[\mathcal{E}_k : E\left\{x_k x_k^T/y_1, \ \ldots, \ y_k\right\}\right] \ . \tag{2.34}$$

The first expectation in (2.34) is known. The second expectation is evaluated by noting that

$$x_k x_k^T = \left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T + \hat{x}_k x_k^T + x_k \hat{x}_k^T - 2\hat{x}_k \hat{x}_k^T \ . \tag{2.35}$$

But, in (2.35), $\hat{x}_k$ is a known function of the data up to y_k so

$$E\, x_k x_k^T = E\left\{\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T / y_1, \ldots, y_k\right\}$$

$$+ \hat{x}_k E\left\{x_k^T / y_1, \ldots, y_k\right\} + E\left\{x_k / y_1, \ldots, y_k\right\} \hat{x}_k^T - \hat{x}_k \hat{x}_k^T = \bar{P}_k + \hat{x}_k \hat{x}_k^T \tag{2.36}$$

Hence, the prediction Eq. (2.34) can be written

$$\hat{x}_{k+1/k} = \Phi_k \hat{x}_k + \left[\mathcal{E}_k : \hat{x}_k \hat{x}_k^T\right] + \left[\mathcal{E}_k : P_k\right] \tag{2.37}$$

This prediction equation differs from the extended Kalman filter version by a bias correction term. An expression for the covariance of this prediction is derived in much the same way. Two additional variables are introduced to simplify the derivation. The first is the estimation error defined by

$$\tilde{x}_k \triangleq x_k - \hat{x}_k \quad . \tag{2.38}$$

The second is the prediction error

$$\tilde{x}_{k+1/k} \triangleq x_{k+1} - \hat{x}_{k+1/k} \quad . \tag{2.39}$$

When (2.7a) and (2.37) are substituted into (2.39), the resulting expression for the prediction error becomes

$$\tilde{x}_{k+1/k} = \Phi_k \tilde{x}_k + \left[\mathcal{E}_k : \left(x_k x_k^T - \hat{x}_k \hat{x}_k^T\right)\right] + G_k q_k - \left[\mathcal{E}_k : P_k\right] \tag{2.40}$$

It is advantageous to simplify the quadratic terms in by completing the square to get

$$x_k x_k^T - \hat{x}_k \hat{x}_k^T = \tilde{x}_k \tilde{x}_k^T + \hat{x}_k \tilde{x}_k^T + \tilde{x}_k \hat{x}_k^T \tag{2.41}$$

In terms of these new variables, the covariance $\bar{P}_{k+1/k}$ is written

$$\bar{P}_{k+1/k} = E\left\{\tilde{x}_{k+1/k}\tilde{x}_{k+1/k}/y_1, \ \dots, \ y_k\right\} \tag{2.42}$$

When (2.40) is substituted into (2.42) and the result expanded many terms arise which either contain terms in δ^2 which can be ignored, or which are uncorrelated (have zero expectation) such as the product of $\tilde{x}$ with $[\delta_k : P_k]$ and the cross terms containing q_k. The only terms which at first glance appear different from the linear case are one involving cubics in $\tilde{x}$, x and $\hat{x}$. If, however, the system (2.7a) and (2.7b) came about by an expansion of a system like (2.1), these are in fact higher order terms and must also be neglected. It is interesting to note that Athans et. al. (Ref. 2) did not notice this in a similar development for a continuous filter without state noise. Their simulation experiments, however, justified the above statements since they found this term to have negligible effect. Hence, (2.42) can be rewritten as

$$\bar{P}_{k+1/k} = \Phi_k \bar{P}_k \Phi_k^T + G_k Q_k G_k^T \tag{2.43}$$

The next step is to determine

$$\hat{x}_{k+1/k+1} = E\left\{x_{k+1}/y_1, \ \dots, \ y_k, y_{k+1}\right\} \tag{2.44}$$

as a function of $\hat{x}_{k+1/k}$. That is, the estimate at step $k+1$ must be updated to reflect the new data available at step $k+1$. In theory, the expectation (2.44) can be determined if the joint conditional probability density

$$p(x_{k+1}, y_{k+1}/y_1, \ \dots, \ y_k) \tag{2.45}$$

is known.

In practice, for all but linear problems, this density tends to be difficult to compute. The first two central moments, however, can be approximated by the same techniques used to obtain the prediction equations.

This is easily done by first computing the conditional expectation of the $(n+m)$ dimensional column vector obtained by augmenting y_{k+1} to the vector x_{k+1}. Thus we can write

$$\begin{bmatrix} \hat{x}_{k+1/k} \\ \hat{y}_{k+1/k} \end{bmatrix} = E\left\{ \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} \Big/ y_1, \; \ldots, \; y_k \right\}$$

$$= \begin{bmatrix} \hat{x}_{k+1/k} \\ H_{k+1}\hat{x}_{k+1/k} + E_{k+1}:\hat{x}_{k+1/k}\hat{x}^T_{k+1/k} + E_{k+1}:P_{k+1/k} \end{bmatrix} \tag{2.46}$$

The second moment is computed from

$$A(k+1) = E\left\{ \begin{bmatrix} x_{k+1} - \hat{x}_{k+1/k} \\ y_{k+1} - \hat{y}_{k+1/k} \end{bmatrix} \begin{bmatrix} x_{k+1} - \hat{x}_{k+1/k} \\ y_{k+1} - \hat{y}_{k+1/k} \end{bmatrix}^T \Big/ y_1, \; \ldots, \; y_k \right\} \tag{2.47}$$

The last equation is conveniently computed in parts. The first part is the $n \times n$ matrix which we can define as

$$A_{xx}(k+1) = E\left\{\left(x_{k+1} - \hat{x}_{k+1/k}\right)\left(x_{k+1} - \hat{x}_{k+1/k}\right)^T / y_1, \; \ldots, \; y_k\right\} = \bar{P}_{k+1/k} \tag{2.48a}$$

The next part is an $m \times n$ matrix defined by

$$A_{yx}(k+1) = E\left\{\left(y_{k+1} - H_{k+1}\hat{x}_{k+1/k} - \left[E_{k+1}:\hat{x}_{k+1/k}\hat{x}^T_{k+1/k} + P_{k+1}\right]\right)\left(x_{k+1} - \hat{x}_{k+1/k}\right)^T / y_1, \; \ldots, \; y_k\right\}$$

$$= E\left\{\left[\nu_{k+1} + H_{k+1}\left(x_{k+1} - \hat{x}_{k+1/k}\right)\right]\left(x_{k+1} - \hat{x}_{k+1/k}\right)^T / y_1, \; \ldots, \; y_k\right\} \tag{2.48b}$$

$$\cong H_{k+1}\bar{P}_{k+1/k}$$

The last part to be determined is the $m \times m$ matrix

$$A_{yx}(k+1) = E\left\{\left(y_{k+1} - H_{k+1}\hat{x}_{k+1/k}\right)\left(y_{k+1} - H_{k+1}\hat{x}_{k+1/k}\right)^T / y_1, \ \ldots, \ y_k\right\} \tag{2.48c}$$

where, as before, cubic terms in x are dropped consistent with the original expansions. The covariance matrix A in Eq. (2.47) is now written

$$A(k+1) = \begin{bmatrix} A_{xx}(k+1) & A_{yx}^T(k+1) \\ A_{yx}(k+1) & A_{yy}(k+1) \end{bmatrix} \tag{2.48d}$$

Although the mean and covariance of the distribution (2.45) are not, in general, sufficient to completely specify the distribution, we circumvent this difficulty by pretending that it is gaussian. This might appear unduly restrictive but, in fact, the solution we will get for (2.44) will be the same one that would be obtained by substituting a linear least square criterion for the conditional mean criterion (e.g., Ref. 17, p. 46). Hence the distribution (2.45) is now written as

$$P(x_{k+1}, y_{k+1}/y_1, \ \ldots, \ y_k) = \frac{1}{(2\pi)^{\frac{n+m}{2}} \left|A(k+1)\right|^{1/2}}$$

$$\cdot \exp\left[-\frac{1}{2}\left(x_{k+1}^T - \hat{x}_{k+1/k}^T, y_{k+1}^T - \hat{y}_{k+1/k}^T\right)\right.$$

$$\left.\cdot A^{-1}(k+1)\left(x_k - \hat{x}_{k+1/k}, y_{k+1} - \hat{y}_{k+1/k}\right)\right] \tag{2.49}$$

The conditional mean and covariance we desire can be directly written down for this case using results given in reference (1). Hence

$$E\left\{x_{k+1}/y_1, \ldots, y_k, y_{k+1}\right\} = \hat{x}_{k+1/k} + A_{xy}A_{yy}^{-1}\left(y_{k+1} - \hat{y}_{k+1/k}\right) \tag{2.50}$$

and

$$E\left\{\left(x_{k+1} - \hat{x}_{k+1/k+1}\right)\left(x_{k+1} - \hat{x}_{k+1/k+1}\right)^T / y_1, y_2, \ldots, y_{k+1}\right\} = A_{xx} - A_{yx}^T A_{yy}^{-1} A_{yx} \quad . \tag{2.51}$$

Aside from some algebraic manipulation to achieve more common forms for the above equations, an approximate forward solution for finding a conditional mean has been obtained.

The recursive equations for computing the approximate conditional mean are now summarized. They are:

$$\hat{x}_{\ell+1/\ell} = \Phi_\ell \hat{x}_{\ell/\ell} + \left[\mathscr{E}_\ell : \hat{x}_{\ell/\ell}\hat{x}_{\ell/\ell}^T\right] + \left[\mathscr{E}_\ell : P_\ell\right] \tag{2.52a}$$

$$H'_{\ell+1} \triangleq H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right] \tag{2.52b}$$

$$\bar{P}_{\ell+1/\ell} = \Phi_\ell \bar{P}_\ell \Phi_\ell^T + G_\ell Q_\ell G_\ell^T \tag{2.52c}$$

$$\bar{P}_{\ell+1}^{-1} = P_{\ell+1/\ell}^{-1} + H'^T_{\ell+1} R_{\ell+1}^{-1} H'_{\ell+1} \tag{2.52d}$$

$$\hat{y}_{\ell+1/\ell} = H_{\ell+1}\hat{x}_{\ell+1/\ell} + \left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\hat{x}_{\ell+1/\ell}^T\right] + \left[E_{\ell+1} : P_{\ell+1/\ell}\right] \tag{2.52e}$$

$$\hat{x}_{\ell+1/\ell+1} = \hat{x}_{\ell+1/\ell} + P_{\ell+1} H'^T_{\ell+1}\left(y_{\ell+1} - \hat{y}_{\ell+1/\ell}\right) \tag{2.52f}$$

This concludes the development of the approximate minimum variance equation.

2.5 An Iterated Extended Kalman Filter

This filter is very similar to the Extended Kalman Filter. However, in the derivation of that filter, a solution was obtained by linearizing about a predicted value defined by Eq. (2.20). For this filter, the minimum to Eq. (2.19) is approximated by relinearizing the residual (2.22) about the estimate found by an extended Kalman filter step and performing the minimization a second time. We begin the derivation of this filter by defining the estimate obtained by using Eqs. (2.30a-g) as $\hat{x}^{(1)}_{\ell+1/\ell+1}$. Corresponding to (2.17), a (hopefully) small excursion is defined by

$$\delta\hat{x}^{(2)}_{\ell+1} \triangleq \hat{x}^{(2)}_{\ell+1/\ell+1} - \hat{x}^{(1)}_{\ell+1/\ell+1} \tag{2.53}$$

so that the expansion of the residual about $\hat{x}^{(1)}_{\ell+1/\ell+1}$ is conveniently written as

$$\begin{aligned} y_{\ell+1} - \hat{y}_{\ell+1} &\cong y_{\ell+1} - H_{\ell+1}\hat{x}^{(1)}_{\ell+1/\ell+1} - \left[E_{\ell+1} : \hat{x}^{(1)}_{\ell+1/\ell+1}\hat{x}^{(1)^T}_{\ell+1/\ell+1}\right] \\ &\quad - \left(H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}^{(1)}_{\ell+1/\ell+1}\right]\right)\delta\hat{x}^{(2)}_{\ell+1/\ell+1} \\ &\triangleq y_{\ell+1} - \hat{y}^{(1)}_{\ell+1} - H^{(2)}_{\ell+1}\delta\hat{x}^{(2)}_{\ell+1/\ell+1} \end{aligned} \tag{2.54}$$

The constraint equation which relates

$$\delta\hat{x}^{(2)}_{\ell/\ell+1} \triangleq \hat{x}^{(2)}_{\ell/\ell+1} - \hat{x}_{\ell/\ell} \tag{2.55}$$

to $\delta\hat{x}^{(1)}_{\ell+1/\ell+1}$ is obtained by differencing

$$\hat{x}^{(2)}_{\ell+1/\ell+1} \triangleq \Phi_\ell\hat{x}^{(2)}_{\ell/\ell+1} + \left[\mathcal{E}_\ell : \hat{x}^{(2)}_{\ell/\ell+1}\hat{x}^{(2)^T}_{\ell/\ell+1}\right] + G_\ell\hat{q}^{(2)}_\ell \tag{2.56}$$

with Eq. (2.30f) [using (2.30a) also] to get

$$\delta\hat{x}^{(2)}_{\ell+1/\ell+1} = \Phi_\ell \delta\hat{x}^{(2)}_{\ell/\ell+1} + G_\ell \hat{q}_\ell + P^{*(1)}_{\ell+1} H^{*(1)^T}_{\ell+1} R^{-1}_{\ell+1}\left(y_{\ell+1} - \hat{y}_{\ell+1/\ell}\right) \quad (2.57)$$

The critical hypothesis in the last equation is that the iterated solution will give estimate changes of the same or higher order than the non-linearities so that the term

$$\left[\mathcal{E}_\ell : \hat{x}^{(2)}_{\ell/\ell+1} \hat{x}^{(2)^T}_{\ell/\ell+1} - \hat{x}_{\ell/\ell} \hat{x}^T_{\ell/\ell}\right]$$

can be ignored. No explicit assumption about the size of $\hat{q}$ is made in this iteration. The object now is to minimize

$$\varphi_{\ell+1} \cong \frac{1}{2} \delta\hat{x}^{(2)^T}_{\ell/\ell+1} \left(P^{*(2)}_\ell\right)^{-1} \delta\hat{x}^{(2)}_{\ell/\ell+1} + \frac{1}{2} \hat{q}^T_\ell Q^{-1}_\ell \hat{q}_\ell$$

$$+ \frac{1}{2} \left(y_{\ell+1} - \hat{y}^{(1)}_{\ell+1} - H^{(2)}_{\ell+1} \delta\hat{x}^{(2)}_{\ell+1/\ell+1}\right)^T R^{-1}_{\ell+1} \left(y_{\ell+1} - \hat{y}^{(1)}_{\ell+1} - H^{(2)}_{\ell+1} \delta\hat{x}^{(2)}_{\ell+1/\ell+1}\right) \quad (2.58)$$

subject to the constraint (2.56). The solution is easily obtained using the techniques of Appendix A. The solution can be put in several forms. For convenience, the entire computational sequence is written below:

$$\hat{x}_{\ell+1/\ell} = \Phi_\ell \hat{x}^{(2)}_{\ell/\ell} + \left[\mathcal{E}_\ell : \hat{x}^{(2)}_\ell \hat{x}^{(2)^T}_\ell\right] \quad (2.59a)$$

$$H^{(1)}_{\ell+1} = H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right] \quad (2.59b)$$

$$P^*_{\ell+1/\ell} = \Phi_\ell P^{*(2)}_\ell \Phi^T_\ell + G_\ell Q_\ell G^T_\ell \quad (2.59c)$$

$$\left(P^{*(1)}_{\ell+1}\right)^{-1} = \left(P^*_{\ell+1/\ell}\right)^{-1} + H^{(1)^T}_{\ell+1} R^{-1}_{\ell+1} H^{(1)}_{\ell+1} \quad (2.59d)$$

$$\hat{y}_{\ell+1/\ell} = H_{\ell+1}\hat{x}_{\ell+1/\ell} + \left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\hat{x}^T_{\ell+1/\ell}\right] \tag{2.59e}$$

$$\hat{x}^{(1)}_{\ell+1/\ell+1} = \hat{x}_{\ell+1/\ell} + P^{*(1)}_{\ell+1} H^{(1)^T}_{\ell+1} R^{-1}_{\ell+1}\left(y_{\ell+1} - \hat{y}_{\ell+1/\ell}\right) \tag{2.59f}$$

$$H^{(2)}_{\ell+1} = H_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}^{(1)}_{\ell+1/\ell+1}\right] \tag{2.59g}$$

$$\left(P^{*(2)}_{\ell+1}\right)^{-1} = \left(P^{*}_{\ell+1/\ell}\right)^{-1} + H^{(2)^T}_{\ell+1} R^{-1}_{\ell+1} H^{(2)}_{\ell+1} \tag{2.59h}$$

$$\hat{y}^{(1)}_{\ell+1} = H_{\ell+1}\hat{x}^{(1)}_{\ell+1/\ell+1} + \left[E_{\ell+1} : \hat{x}^{(1)}_{\ell+1/\ell+1}\hat{x}^{(1)^T}_{\ell+1/\ell+1}\right] \tag{2.59i}$$

$$\begin{aligned}\hat{x}^{(2)}_{\ell+1/\ell+1} = \hat{x}_{\ell+1/\ell} + P^{*(2)}_{\ell+1} H^{(2)^T}_{\ell+1} R^{-1}_{\ell+1}\Big[y_{\ell+1} &- \hat{y}^{(1)}_{\ell+1} \\ &- H^{(2)}_{\ell+1}\left(\hat{x}^{(1)}_{\ell+1/\ell+1} - \hat{x}_{\ell+1/\ell}\right)\Big]\end{aligned} \tag{2.59j}$$

This ends the development of the four types of data processing schemes which will be analyzed in the following chapter.

3. APPROXIMATE EQUATIONS FOR THE MEAN AND COVARIANCE OF DATA PROCESSING ALGORITHMS

Introduction

This chapter analyzes the filtering algorithms presented in Chapter 2. The object of the analysis is to find the bias and covariance of each estimation scheme in terms of equations which can be easily computed without using Monte-Carlo procedures. The equations to be developed here are of approximately the same complexity as the computations required to compute covariances in a Kalman filtering algorithm applied to a linear problem. Furthermore, the equations are all in terms of quantities which are computed from a related linear problem and are thus subject to some direct general interpretation.

The four sections deal with the least square, extended Kalman, approximate minimum variance and iterated extended Kalman algorithms, respectively.

3.1 Mean and Covariance of the Error in the Least Square Algorithm

This algorithm is represented by Eqs. (2.13) through (2.15). In order to obtain an explicit solution for this (TPBVP), the equations are first solved ignoring all the small terms containing ϵ and E to get a linear problem. The solution to this set of equations is then substituted into the small terms of (2.12) through (2.15) which are then regarded as driving terms.

The resulting linear smoothing problem is solved in Appendix A. In the sequel, the linear smoothed solutions will simply be defined by unsuperscripted quantities such as $\hat{x}_{k/N}$ for a linear smoothed estimate of x_k given N data points and $P_{k/N}$ for its covariance. The least square estimate is denoted by a superscript (LS). With these conventions, Eqs. (2.12) through (2.15) become:

$$\lambda_{k-1}^{LS} = \Phi_k^T \lambda_k^{LS} + H_k^T R_k^{-1} \left(y_k - H_{k+1} \hat{x}_{k+1}^{LS} \right)$$

$$+ 2 \left[\mathcal{E}_k : \hat{x}_{k/N} \right]^T \lambda_k + 2 \left[E_k : \hat{x}_{k/N} \right]^T R_k^{-1} \left(y_k - h_k \hat{x}_{k/N} \right)$$

$$- H_k^T R_k^{-1} \left[E_k : \hat{x}_{k/N} \hat{x}_{k/N}^T \right] + \text{H.O.T.} \tag{3.1a}$$

$$\lambda_N^{LS} = 0 \tag{3.1b}$$

$$- P_0^{-1} \left(x_0 - \hat{x}_0^{LS} \right) + \Phi_0^T \lambda_0^{LS} + 2 \left[\mathcal{E}_0 : \hat{x}_{0/0} \right]^T \lambda_0 + \text{H.O.T.} = 0 \tag{3.1c}$$

$$\hat{x}_{k+1}^{LS} = \Phi_k \hat{x}_k^{LS} - G_k Q_k G_k^T \lambda_k^{LS} + \left[\mathcal{E}_k : \hat{x}_{k/N} \hat{x}_{k/N} \right] + \text{H.O.T.} \tag{3.1d}$$

In these equations all the small terms are regarded as known terms such as y_k or the deterministic driving term u_k included in the solution in Appendix A. In order to avoid the algebra inherent in solving these equations directly, a correspondence is made between the data and driving terms occurring in the linear problem and that in Eqs. (3.1a) through (3.1d). To do this we define two new variables:

$$u_k' \triangleq \left[\mathcal{E}_k : \hat{x}_{k/N} \hat{x}_{k/N}^T \right] \tag{3.2}$$

$$Y_k' = H_k^T R_k^{-1} y_k + 2 \left[\mathcal{E}_k : \hat{x}_{k/N} \right]^T \lambda_k + 2 \left[E_k : \hat{x}_{k/N} \right]^T R_k^{-1} \left(y_k - H_k \hat{x}_{k/N} \right)$$

$$- H_k^T R_k^{-1} \left[E_k : \hat{x}_{k/N} \hat{x}_{k/N}^T \right] \tag{3.3}$$

In the linear relations of Appendix A data always occurs as a driving term of the form

$$Y_k = H_k^T R_k^{-1} y_k \tag{3.4}$$

Hence, in the solutions of Appendix A, wherever (3.4) appears, (3.3) is substituted and wherever u_k appears, (3.2) is used instead. The correspondence is completed by making the boundary condition (3.1c) look like Eq. (A.8) by defining a pseudo a priori mean

$$\bar{x}_0' = \bar{x}_0 = 2P_0\left[\delta_0 : \hat{x}_{0/N}\right]^T \lambda_0 \tag{3.5}$$

The solutions to equations (3.1a) through (3.1d) are now written down in terms of the well known forward difference equations which must be solved up to point N to get $\hat{x}^{LS}_{N/N}$. It is important to note that the intermediate points in the solution of these equations do not represent approximations to the least square solution at points $k < N$ since the driving terms are based on a zeroeth order approximation using N data points. As a matter of convenience, however, the equations are written in the same notation employed for the usual filtering solutions:

$$\hat{x}^{LS}_{\ell+1/\ell} = \Phi_\ell \hat{x}^{LS}_\ell + u'_\ell \tag{3.6a}$$

$$P_{\ell+1/\ell} = \Phi_\ell P_\ell \Phi_\ell^T + G_\ell Q_\ell G_\ell^T \tag{3.6b}$$

$$P^{-1}_{\ell+1} = P^{-1}_{\ell+1/\ell} + H^T_{\ell+1} R^{-1}_{\ell+1} H_{\ell+1} \tag{3.6c}$$

$$\hat{x}^{LS}_{\ell+1} = \hat{x}^{LS}_{\ell+1/\ell} + P_{\ell+1}\left(Y'_{\ell+1} - H^T_{\ell+1} R^{-1}_{\ell+1} H_{\ell+1} \hat{x}^{LS}_{\ell+1/\ell}\right) \tag{3.6d}$$

$$\hat{x}^{LS}_0 = \bar{x}_0 - 2P_0\left[\delta_0 : \hat{x}_{0/N}\right]^T \lambda_0 \tag{3.6e}$$

In order to obtain equations for the mean and covariance of the least square estimate, the Eqs. (3.6a) through (3.6e) will now be rewritten in error equation form. A good deal of algebra will be avoided by using well known probabilistic relations for the linear estimates occurring in the driving terms. The error notation is described in Eqs. (2.38) and (2.39).

When Eq. (3.6a) is substituted into (3.6d) and the result subtracted from the system Eq. (2.7a), we get:

$$\tilde{x}^{LS}_{\ell+1} = \Phi_\ell \tilde{x}^{LS}_\ell + G_\ell q_\ell + \left[\mathcal{E}_\ell : \left(x_\ell x^T_\ell - \hat{x}_{\ell/N}\hat{x}^T_{\ell/N}\right)\right]$$

$$+ P_{\ell+1}\left[Y'_{\ell+1} - H^T_{\ell+1}R^{-1}_{\ell+1}H_{\ell+1}\left(\Phi_\ell \hat{x}^{LS}_\ell + u'_\ell\right)\right] \tag{3.7}$$

The last term in (3.7) is expanded further by recalling Eq. (3.3) and the relation

$$y_{\ell+1} = H_{\ell+1}\left(\Phi_\ell x_\ell + \left[\mathcal{E}_\ell : x_\ell x^T_\ell\right] + G_\ell q_\ell\right)$$

$$+ \left[E_{\ell+1} : x_{\ell+1}x^T_{\ell+1}\right] + \nu_{\ell+1} \tag{3.8}$$

When these terms are substituted into (3.7) and grouped appropriately we obtain the result

$$\tilde{x}^{LS}_{\ell+1} = \Phi_\ell \tilde{x}^{LS}_\ell + G_\ell q_\ell + \left[\mathcal{E}_\ell : \left(x_\ell x^T_\ell - \hat{x}_{\ell/N}\hat{x}^T_{\ell/N}\right)\right]$$

$$- P_{\ell+1}H^T_{\ell+1}R^{-1}_{\ell+1}\left(H_{\ell+1}\left(\Phi_\ell \tilde{x}^{LS}_\ell + G_\ell q_\ell\right) + \left[\mathcal{E}_\ell : \left(x_\ell x^T_\ell - \hat{x}_{\ell/N}\hat{x}^T_{\ell/N}\right)\right] + \nu_{\ell+1}\right)$$

$$- P_{\ell+1}H^T_{\ell+1}R^{-1}_{\ell+1}\left[E_{\ell+1} : \left(\hat{x}_{\ell+1/N}\hat{x}_{\ell+1/N} - x_{\ell+1}x^T_{\ell+1}\right)\right]$$

$$+ 2P_{\ell+1}\left[\mathcal{E}_{\ell+1} : \hat{x}_{\ell+1/N}\right]^T \lambda_{\ell+1} + 2P_{\ell+1}\left[E_{\ell+1} : \hat{x}_{\ell+1/N}\right]^T R^{-1}_{\ell+1}$$

$$\cdot \left(H_{\ell+1}\hat{x}_{\ell+1/N} + \nu_{\ell+1}\right) \tag{3.9}$$

The mean of the least square algorithm error is determined by finding the expected value of the difference Eq. (3.9) over the random variables ν_k and q_k the initial error. The expectation of the terms which do not involve either E_k or $\mathcal{E}_k$ is obvious. The expectation of the terms containing the small non-linearities is found using the linear theory developed in the Appendices. The first of these requires the evaluation of

$$E\left\{x_\ell x_\ell^T - \hat{x}_{\ell/N}\hat{x}^T_{\ell/N}\right\} = E\left\{\hat{x}_{\ell/N}\hat{x}^T_{\ell/N} + \hat{x}_{\ell/N}\tilde{x}^T_{\ell/N} + \tilde{x}_{\ell/N}\hat{x}^T_{\ell/N}\right\} \tag{3.10}$$

where the last two terms on the right side of (3.10) have zero expectation since Appendix B shows that the linear estimate and its error are uncorrelated. Hence, we see immediately that

$$E\left\{x_\ell x_\ell^T - \hat{x}_{\ell/N}\hat{x}^T_{\ell+N}\right\} = P_{\ell/N} \tag{3.11}$$

The term containing the adjoint $\lambda_{\ell+1}$ will consist entirely of linear combinations of products occurring in the matrix representing the outer product between $\hat{x}_{\ell+1/N}$ and $\lambda_{\ell+1}$. However, using the relation (A.45), we can show that the expected value of the outer product matrix is zero. In fact, adding and subtracting x_ℓ to (A.45) immediately gives

$$\lambda_{\ell+1} = \Phi^{-T}_{\ell+1}P^{-1}_{\ell+1}\left(\tilde{x}_{\ell+1/N} - \tilde{x}_{\ell+1}\right) \tag{3.12}$$

and we have shown in Appendix B the well known fact that

$$E\left\{\hat{x}_{\ell+1/N}\tilde{x}^T_{\ell+1/N}\right\} = 0 \tag{3.13a}$$

and the lesser known fact that

$$E\left\{\hat{x}_{\ell+1/N}\tilde{x}^T_{\ell+1}\right\} = 0 \tag{3.13b}$$

so the result is demonstrated.

The next term of interest involves linear combinations of terms occuring in the outer product

$$E\left\{\hat{x}_{\ell+1/N}\tilde{x}^T_{\ell+1/N}\right\} = 0$$

and also a term

$$E\left\{2P_{\ell+1}\left[E_{\ell+1}:\hat{x}_{\ell+1/N}\right]^T R^{-1}_{\ell+1}\,\nu_{\ell+1}\right\} \tag{3.14a}$$

which will contain linear combinations of the quantities occuring in the outer product between $\hat{x}_{\ell+1/N}$ and $\nu_{\ell+1}$. In Appendix B, however, it is shown that

$$E\left\{\hat{x}_{\ell+1/N}V^T_{\ell+1}\right\} = P_{\ell+1/N}H^T_{\ell+1} \ . \tag{3.14b}$$

With the use of (3.14b), the term (3.14a) becomes

$$2P_{\ell+1}\left[E^T_{\ell+1}:\left(P_{\ell+1/N}H^T_{\ell+1}R^{-1}_{\ell+1}\right)\right]^* \tag{3.14c}$$

The remaining term is analogous to (3.10). Hence, the desired expectation can immediately be written:

---

*In explicit summation notation this term is written as

$$2\sum_{\beta,\gamma,\delta,\rho,\xi} P_{\alpha\beta}(\ell+1)\ E^{(\ell+1)}_{\gamma\beta\delta}\ P_{\gamma\rho}(\ell+1/N)\ H^{(\ell+1)}_{\xi\rho}\ R^{-1}_{\xi\delta}(\ell+1)$$

$$E\left\{\tilde{x}^{LS}_{\ell+1}\right\} = \tilde{\Phi}_\ell \, E\left\{\tilde{x}^{LS}_{\ell}\right\} + \left(I - P_{\ell+1}H^T_{\ell+1}R^{-1}_{\ell+1}H_{\ell+1}\right)\left[\mathcal{E}_\ell : P_{\ell/N}\right]$$

$$- P_{\ell+1}\left(2\left[E^T_{\ell+1} : \left(P_{\ell+1/N}H^T_{\ell+1}R_{\ell+1}\right)\right] + H^T_{\ell+1}R^{-1}_{\ell+1}\left[E_{\ell+1} : P_{\ell+1/N}\right]\right) \tag{3.15}$$

where the error transition matrix is defined by

$$\tilde{\Phi}_\ell = \left(I - P_{\ell+1}H^T_{\ell+1}R^{-1}_{\ell+1}H_{\ell+1}\right)\Phi_\ell \; . \tag{3.16}$$

Providing $\bar{x}_0$ represents the true a priori mean of x_0, the algorithm (3.15) is started with the initial condition

$$E\left\{\tilde{x}^{LS}_0\right\} = 0 \; . \tag{3.17}$$

In order to obtain an estimate of the covariance of $\tilde{x}^{LS}_N$ without being ensnared in an intolerable amount of algebra the solution to (3.9) is written explicitly in summation notation, the initial condition (3.6e) substituted into the result, and the first order terms grouped to get

$$\tilde{x}^{LS}_N = \tilde{x}_N + 2\sum_{i=0}^{n} \tilde{\Phi}_{N/i}P_i\left[\mathcal{E}_i : \hat{x}_{i/N}\right]\Phi^T_i P^{-1}_i\left(\tilde{x}_{i/N} - \tilde{x}_i\right)$$

$$+ \sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}\left(I - P_{i+1}S_{i+1}\right)\left[\mathcal{E}_i : x_i x^T_i - \hat{x}_{i/N}\hat{x}^T_{i/N}\right]$$

$$+ 2\sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}P_{i+1}\left[E_{i+1} : \hat{x}_{i+1/N}\right]^T R^{-1}_{i+1}\left(H_{i+1}\tilde{x}_{i+1/N} + \nu_{i+1}\right)$$

$$+ \sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}P_{i+1}H^T_{i+1}R^{-1}_{i+1}\left[E_{i+1} : \left(x_{i+1}x^T_{i+1} - \hat{x}_{i+1/N}\hat{x}^T_{i+1/N}\right)\right] \tag{3.18}$$

The covariance of the error in $\hat{x}_N^{LS}$ based on an ensemble of cases having various initial conditions x_0 and various noise sequences q_k and p_k will, in general, differ from a conditional covariance based on a particular sequence of data. In the linear case, these covariances are the same. The ensemble covariance is defined by

$$P_N^{LS} \triangleq E\left\{\left(\tilde{x}_N^{LS} - E\left\{\tilde{x}_N^{LS}\right\}\right)\left(\tilde{x}_N^{LS} - E\left\{\tilde{x}_N^{LS}\right\}\right)^T\right\} \tag{3.19a}$$

$$= E\left\{\tilde{x}_N^{LS}\tilde{x}_N^{LS}\right\} - E\left\{\tilde{x}_N^{LS}\right\} E\left\{\tilde{x}_N^{LS}\right\}^T \tag{3.19b}$$

The first term on the right side of (3.19b) is found by squaring (3.18) and taking an expectation. The second term will be entirely higher order and so will be ignored. At this point some additional assumptions are made. They are: a) the random variable x_0 is drawn from a distribution which is symmetric about the <u>a priori</u> mean $\bar{x}_0$ and is entirely independent of the q_k and ν_k; and b) the sequences q_k v_k are drawn from independent zero mean symmetric distributions. From the results given in Appendix A, it is clear that such terms as $\tilde{x}_{j/N}$ and $\tilde{x}_j$ can be written as linear combinations of $\tilde{x}_0$, and the q_k ν_k sequences. Together with assumptions a) and b) above, this leads to the conclusion that, when (3.18) is squared, all cubic terms in $\tilde{x}$, q and ν will yield zero expectation. Terms involving products of $\mathcal{E}$ and E are by hypothesis negligible. Furthermore, the terms in Eq. (3.18) involving the difference of the squares of the state and its estimate are replaced by the same relation used in (3.10) so that

$$P_N^{LS} \approx P_N + A_N + A_N^T \tag{3.20}$$

where

$$A_N = E\left\{\sum_{i=0}^{N-1} \tilde{\Phi}_{N/i} P_i \left[\mathcal{E}_i : \hat{x}_{i/N}\right] \Phi_N^{-T} P_i^{-1}\left(\tilde{x}_{i/N} - \tilde{x}_i\right) \tilde{x}_N^T\right\}$$

$$+ E\left\{\sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}(I - P_{i+1}S_{i+1})\left[\mathcal{E}_i : \hat{x}_{i/N}\tilde{x}^T_{i/N} + \tilde{x}_{i/N}\hat{x}^T_{i/N}\right]\tilde{x}^T_N\right\}$$

$$+ E\ 2\left\{\sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}P_{i+1}\left[E_{i+1} : \hat{x}_{i+1/N}\right]^T R^{-1}_{i+1}\left(H_{i+1}\tilde{x}_{i+1} + \nu_{i+1}\right)\tilde{x}^T_N\right\}$$

$$+ E\left\{\sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}P_{i+1}H^T_{i+1}R^{-1}_{i+1}\left[E_{i+1} : \hat{x}_{i+1/N}\tilde{x}^T_{i+1/N} + \tilde{x}_{i+1/N}\hat{x}^T_{i+1/N}\right]\tilde{x}^T_N\right\}$$

(3.21)

The term $\hat{x}_{i/N}$ in (3.21) can be written as

$$\hat{x}_{i/N} = x_i - \tilde{x}_{i/N} \tag{3.22}$$

But $\tilde{x}_{i/N}$ is simply a linear combination of $\tilde{x}_0$ and the q_k and ν_k, and second order terms so that by the hypothesis above it will give no contribution to the expectations in (3.21). By similar reasoning, x_i can be written as a linear combination of $\Phi_{i/0}x_0$ plus terms with q_k and second order effects. x_0 can, in turn be written as $\bar{x}_0 + \tilde{x}_0$ from which we see immediately that all the $\hat{x}_{i/N}$ terms in (3.22) can be replaced by $\Phi_{N/i}\bar{x}_0$ which is deterministic in character. Hence, the expectations in (3.22) need only be taken over the remaining products of two errors in each term.

From the results in Appendix B,

$$E\left\{\tilde{x}_{i/N}\tilde{x}^T_N\right\} = E\left\{\tilde{x}_i\tilde{x}_N\right\} = P_i\tilde{\Phi}^T_{N/i} \tag{3.23}$$

and

$$E\ \left\{\nu_{i+1}\tilde{x}^T_N\right\} = -H_{i+1}P_{i+1}\tilde{\Phi}^T_{N/i+1} \tag{3.24}$$

so that the first and third terms in (3.21) have zero expectation. The remaining terms are written as

$$A_N = 2 \sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}(I - P_{i+1}S_{i+1})\left[\mathcal{E}_i : \Phi_{i/0}\bar{x}_0\right]P_i\tilde{\Phi}^T_{N/i}$$

$$+ 2 \sum_{i=0}^{N-1} \tilde{\Phi}_{N/i+1}P_{i+1}H^T_{i+1}\left[E_{i+1} : \Phi_{i+1/0}\bar{x}_0\right]P_{i+1}\tilde{\Phi}^T_{N/i+1} \quad . \tag{3.25}$$

Equations (3.15), (3.20) and (3.25) complete the derivations of approximations for the approximate mean and covariance of the errors in a least square estimation.

3.2 Mean and Covariance of the Error in the Extended Kalman Filter (EK)

The first step in analyzing this algorithm is to use Eqs. (2.30a) through (2.30g) to obtain the state error equation:

$$\tilde{x}^{EK}_{\ell+1} = \Phi_\ell\tilde{x}^{EK}_\ell + \left[\mathcal{E}_\ell : x_\ell x^T_\ell - \hat{x}^{EK}_\ell\hat{x}^{EK^T}_\ell\right] + G_\ell q_\ell$$

$$- P^*_{\ell+1}\left(H^T_{\ell+1} + 2\left[E_{\ell+1} : \Phi_\ell\hat{x}^{EK}_\ell\right]\right)R^{-1}_{\ell+1}$$

$$\cdot \left\{\nu_{\ell+1} + H_{\ell+1}\left(\Phi_\ell\tilde{x}^{EK}_\ell + \left[\mathcal{E}_\ell : x_\ell x^T_\ell - \hat{x}^{EK}_\ell\hat{x}^{EK^T}_\ell\right] + G_\ell q_\ell\right)\right.$$

$$\left. + \left[E_{\ell+1} : \left(x_{\ell+1}x^T_{\ell+1} - \hat{x}^{EK}_{\ell+1}\hat{x}^{EK}_{\ell+1}\right)\right]\right\} + HOT \tag{3.26}$$

The term $P^*_{\ell+1}$ is also a function of the $\hat{x}^{EK}$ quantity and is updated according to (2.30c) and the equation

$$P^*_{\ell+1} = P^*_{\ell+1/\ell} - P^*_{\ell+1/\ell}H'^T_{\ell+1}\left(H'_{\ell+1}P^*_{\ell+1/\ell}H'^T_{\ell+1} + R_{\ell+1}\right)^{-1}H'_{\ell+1}P^*_{\ell+1} \tag{3.27}$$

The last equation can be put in a form where higher order terms can be separated from lower order ones by expanding the matrix inverse term.

To simplify the algebra, a new variable $M_{\ell+1}$ is defined by

$$M^*_{\ell+1} = H_{\ell+1}P^*_{\ell+1}H_{\ell+1} + R_{\ell+1} \tag{3.28a}$$

$$M_{\ell+1} = H_{\ell+1}P_{\ell+1}H_{\ell+1} + R_{\ell+1} \tag{3.28b}$$

The matrix inverse is now written as

$$\left(H'_{\ell+1}P^*_{\ell+1/\ell}H'^T_{\ell+1} + R_{\ell+1}\right)^{-1} = M^{*-1}_{\ell+1}\Bigg(I + 2\left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]P^*_{\ell+1/\ell}H^T_{\ell+1}M^{*-1}_{\ell+1}$$

$$+ 2H_{\ell+1}P^*_{\ell+1/\ell}\left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]^T M^{*-1}_{\ell+1} + HOT\Bigg)^{-1}$$

$$\cong M^{*-1}_{\ell+1} - 2M^{*-1}_{\ell+1}\left(\left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]P^*_{\ell+1/\ell}H^T_{\ell+1} + H_{\ell+1}P^*_{\ell+1/\ell}\left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]\right)M^{*-1}_{\ell+1} \tag{3.29}$$

When (3.29) is substituted in (3.27) and only first order terms retained it can be seen that

$$P^*_{\ell+1} = P^*_{\ell+1/\ell} - P^*_{\ell+1/\ell}H^T_{\ell+1}M^{*-1}_{\ell+1}H_{\ell+1}P^*_{\ell+1/\ell}$$

$$+ 2P^*_{\ell+1/\ell}H^T_{\ell+1}M^{*-1}_{\ell+1}\left(\left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]P^*_{\ell+1/\ell}H^T_{\ell+1} + H_{\ell+1}P^*_{\ell+1/\ell}\left[E_{\ell+1/\ell}:\hat{x}^{EK}_{\ell+1/\ell}\right]^T\right)$$

$$\cdot M^{*-1}_{\ell+1}H_{\ell+1}P^*_{\ell+1/\ell} - 2P^*_{\ell+1/\ell}\Bigg(H^T_{\ell+1}M^{*-1}_{\ell+1}\left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]$$

$$+ \left[E_{\ell+1}:\hat{x}^{EK}_{\ell+1/\ell}\right]M^{*-1}_{\ell+1}H_{\ell+1}\Bigg)P^*_{\ell+1/\ell} + H.O.T. \tag{3.30}$$

Equation (3.30) and (3.26) show that the non-linear equations for the extended Kalman filter look like the linear equations with higher order terms added. An approximate solution is then obtained by getting a zeroeth order solution which ignores all the non-linearities and substituting this linear recursive solution back into the small terms of (3.26) and (3.30). When this is done, the update equations for the P^* terms becomes

$$P^*_{\ell+1/\ell} = \Phi_\ell P^*_\ell \Phi^T_\ell + G_\ell Q_\ell G^T_\ell \tag{3.31a}$$

$$P^*_{\ell+1} = P^*_{\ell+1/\ell} - P^*_{\ell+1/\ell} H^T_{\ell+1} M^{*-1}_{\ell+1} H_{\ell+1} P^*_{\ell+1/\ell} + \Gamma_{\ell+1} \tag{3.31b}$$

where, after some manipulation,

$$\begin{aligned}\Gamma_{\ell+1} &= -2P_{\ell+1/\ell} H^T_{\ell+1} M^{-1}_{\ell+1} \left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right] P_{\ell+1} \\ &\quad -2P_{\ell+1}\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right]^T M^{-1}_{\ell+1} H_{\ell+1} P_{\ell+1/\ell}\end{aligned} \tag{3.31c}$$

Equations (3.31a) through (3.31c) can be written in error equation form by subtracting them from the linear update equations for the covariance to get

$$\Delta P^{EK}_{\ell+1/\ell} = P_{\ell+1/\ell} - P^*_{\ell+1/\ell} = \Phi_\ell \Delta P^{EK}_\ell \Phi^T_\ell \tag{3.32a}$$

and

$$\begin{aligned}\Delta P^{EK}_{\ell+1} &= \Delta P^{EK}_{\ell+1/\ell} + P_{\ell+1} H^T_{\ell+1} M^{-1}_{\ell+1} \Delta P^{EK}_{\ell+1/\ell} H^T_{\ell+1} M^{-1}_{\ell+1} H_{\ell+1} P_{\ell+1/\ell} \\ &\quad - \Delta P^{EK}_{\ell+1/\ell} H^T_{\ell+1} M^{-1}_{\ell+1} H_{\ell+1} P_{\ell+1/\ell} - P_{\ell+1/\ell} H^T_{\ell+1} M^{-1}_{\ell+1} H_{\ell+1} \Delta P^{EK}_{\ell+1/\ell} \\ &\quad - \Gamma_{\ell+1} + \text{terms of order } (\Delta P^{EK})^2 \\ &= \tilde{\Phi}_{\ell+1/\ell} \Delta P^{EK}_\ell \tilde{\Phi}_{\ell+1/\ell} - \Gamma_{\ell+1}\end{aligned} \tag{3.32b}$$

with

$$\Delta P_0^{EK} = 0 \tag{3.32c}$$

The solution to (3.32b) is written in summation form as

$$\Delta P_{\ell+1}^{EK} = - \sum_{i=1}^{\ell+1} \tilde{\Phi}_{\ell+1/i} \; \Gamma_i \; \tilde{\Phi}_{\ell+1/i} \tag{3.33}$$

The solution to (3.26) is now written as

$$\begin{aligned} \tilde{x}_{\ell+1}^{EK} = \tilde{\Phi}_{\ell+1/\ell}\tilde{x}_{\ell}^{EK} &+ \left(I - P_{\ell+1}H_{\ell+1}^{T}R_{\ell+1}^{-1}H_{\ell+1}\right)\left(G_{\ell}q_{\ell} + \left[\mathcal{E}_{\ell}:x_{\ell}x_{\ell}^{T} - \hat{x}_{\ell}\hat{x}_{\ell}^{T}\right]\right) \\ &- P_{\ell+1}H_{\ell+1}^{T}R_{\ell+1}^{-1}\left\{\nu_{\ell+1}+\left[E_{\ell+1}:\left(x_{\ell+1}x_{\ell+1}^{T} - \hat{x}_{\ell+1/\ell}\hat{x}_{\ell+1/\ell}^{T}\right)\right]\right\} \\ &- 2P_{\ell+1}\left[E_{\ell+1}:\hat{x}_{\ell+1/\ell}\right]^{T} R_{\ell+1}^{-1}\left(\nu_{\ell+1} + H_{\ell+1}\tilde{x}_{\ell+1/\ell}\right) \\ &- \Delta P_{\ell+1}^{EK}H_{\ell+1}^{T}R_{\ell+1}^{-1}\left(\nu_{\ell+1} + H_{\ell+1}\tilde{x}_{\ell+1/\ell}\right) \end{aligned} \tag{3.34}$$

We now take the expectation of both sides of (3.34). Since the term $\Delta P_{\ell+1}^{EK}$ consists of a summation of terms involving the linear filtered estimates up to $\hat{x}_{\ell+1/\ell}$, we know it will be orthogonal to $\nu_{\ell+1}$, $\tilde{x}_{\ell+1/\ell}$ and q_{ℓ}. The other terms follow from Appendix B. Hence,

$$\begin{aligned} E\left\{\tilde{x}_{\ell+1}^{EK}\right\} = \tilde{\Phi}_{\ell+1/\ell}\, E\left\{\tilde{x}_{\ell}^{EK}\right\} &+ \left(I - P_{\ell+1}H_{\ell+1}^{T}R_{\ell+1}^{-1}H_{\ell+1}\right)\left[\mathcal{E}_{\ell}:P_{\ell}\right] \\ &- P_{\ell+1}H_{\ell+1}^{T}R_{\ell+1}^{-1}\left[E_{\ell+1}:P_{\ell+1/\ell}\right] \end{aligned} \tag{3.35}$$

The covariance of the EK filter is computed much like the least square filter and turns out to be exactly the same as that given in Eqs. (3.20) and (3.25) for that case.

3.3 Mean and Covariance of the Error in the Approximate Minimum Variance Filter (AM)

As with the previous two algorithms, an error equation for this filter is found by differencing Eq. (2.52) with the state equation to get

$$\tilde{x}^{AM}_{\ell+1} = \Phi_\ell \tilde{x}^{AM}_\ell + \left[\mathcal{E}_\ell : x_\ell x_\ell^T - \hat{x}^{AM}_\ell \hat{x}^{AM^T}_\ell\right] - \left[\mathcal{E}_\ell : P^*_{\ell/\ell}\right] + G_\ell q_\ell$$
$$- P^*_{\ell+1}\left(H^T_{\ell+1} + 2\left[E_{\ell+1} : \hat{x}^{AM}_{\ell+1/\ell}\right]\right) R^{-1}_{\ell+1}$$
$$\cdot \left(\nu_{\ell+1} + H_{\ell+1}\Phi_\ell \tilde{x}^{AM}_\ell + H_{\ell+1} G_\ell q_\ell + \left[E_{\ell+1} : x_{\ell+1} x^T_{\ell+1} - \hat{x}^{AM}_{\ell+1/\ell} \hat{x}^{AM^T}_{\ell+1/\ell}\right]\right.$$
$$\left. + H_{\ell+1}\left[\mathcal{E}_\ell : x_\ell x_\ell^T - \hat{x}^{AM}_\ell \hat{x}^{AM^T}_\ell\right] - H_{\ell+1}\left[\mathcal{E}_\ell : P_\ell\right] - \left[E_{\ell+1} : P_{\ell+1/\ell}\right]\right) \tag{3.36}$$

Since $P^*_{\ell+1}$ obeys the same update equation as the extended Kalman filter with $\hat{x}^{EK}$ replaced by $\hat{x}^{AM}$ we can immediately write down the approximate solutions to the above equations by first ignoring non-linearities and then resolving the equations with linear solutions substituted into the terms which are small. When this is done, the approximate solution to (3.36) becomes

$$\tilde{x}^{AM}_{\ell+1} = \tilde{\Phi}_\ell \tilde{x}^{AM}_\ell + \left(I - P_{\ell+1} S_{\ell+1}\right)\left(G_\ell q_\ell + \left[\mathcal{E}_\ell : x_\ell x_\ell^T - \hat{x}_\ell \hat{x}_\ell^T\right] - \left[\mathcal{E}_\ell : P_\ell\right]\right)$$
$$- P_{\ell+1} H^T_{\ell+1} R^{-1}_{\ell+1}\left(\nu_{\ell+1} + \left[E_{\ell+1} : x_{\ell+1} x^T_{\ell+1} - \hat{x}_{\ell+1/\ell} \hat{x}^T_{\ell+1/\ell}\right]\right.$$
$$\left. - \left[E_{\ell+1} : P_{\ell+1/\ell}\right]\right) + \Delta P_{\ell+1} H^T_{\ell+1} R^{-1}_{\ell+1}\left(\nu_{\ell+1} + H_{\ell+1}\tilde{x}_{\ell+1/\ell}\right)$$
$$- 2P_{\ell+1}\left[E_{\ell+1} : \hat{x}_{\ell+1/\ell}\right]^T R^{-1}_{\ell+1}\left(\nu_{\ell+1} + H_{\ell+1}\tilde{x}_{\ell+1/\ell}\right) \tag{3.37a}$$

$$\tilde{x}_0^{AM} = 0 \tag{3.37b}$$

Hence, we see that

$$E\left\{\tilde{x}_N^{AM}\right\} = 0 \tag{3.38}$$

When the covariance computation is carried out, it is again found to be the same as that for the least square and extended Kalman algorithms.

3.4 Mean and Covariance of the Error in the Iterated Extended Kalman Filter (IT)

In order to obtain the error equations for the iterated filter it is first necessary to write Eqs. (2.59a-j) in a single step form. When this is done and second order terms discarded it is immediately seen that the iterative step produces only changes of order E from the first step. Hence, in all the small terms involving $\hat{x}^{(1)}$ or $\hat{x}^{(2)}$ it is irrelevant which one is used. This reasoning leads to the analytically useful single step form:

$$\hat{x}^{(IT)}_{k+1/k} = \Phi_k \hat{x}_k^{IT} + \left[\mathcal{E}_k : \hat{x}_k^{IT} \hat{x}_k^{IT^T}\right] \tag{3.39a}$$

$$\hat{y}^{IT}_{k+1/k} = H_{k+1} \hat{x}^{IT}_{k+1/k} + \left[E_{k+1} : \hat{x}^{IT}_{k+1/k} \hat{x}^{IT^T}_{k+1/k}\right] \tag{3.39b}$$

$$\begin{aligned}\hat{x}^{IT}_{k+1} &= \hat{x}^{IT}_{k+1/k} + P^{IT}_{k+1} H^{IT^T}_{k+1} R^{-1}_{k+1} \Big(y_{k+1} - H_{k+1} \hat{x}^{IT}_{k+1/k} \\ &\quad - \left[E_{k+1} : \hat{x}^{IT}_{k+1} \hat{x}^{IT^T}_{k+1}\right] + 2\left[E_{k+1} : \hat{x}^{IT}_{k+1}\right] P_{k+1} H_{k+1} R^{-1}_{k+1} \\ &\quad \cdot \left(y_{k+1} - \hat{y}^{IT}_{k+1/k}\right)\Big) + \text{H O T}\end{aligned} \tag{3.39c}$$

$$H^{IT}_{k+1} \triangleq H_{k+1} + 2\left[E_{k+1} : \hat{x}^{IT}_{k+1}\right] \tag{3.39d}$$

$$\left(P_{k+1}^{IT}\right)^{-1} = \left(P_{k+1/k}^{IT}\right)^{-1} + H_{k+1}^{IT^T} R_{k+1}^{-1} H_{k+1}^{IT} \tag{3.39e}$$

In (3.39c) the P_{k+1} appears since all the P type terms can be written as the linear P plus small corrections which produce only higher order effects in the relevant term. The error equation for the iterated filter can now be written as

$$\begin{aligned}
\tilde{x}_{k+1}^{IT} = \Phi_k \tilde{x}_k^{IT} &+ \left[\mathcal{E}_k : x_k x_k^T - \hat{x}_k^{IT^T} \hat{x}_k^{IT^T}\right] + G_k q_k \\
&+ \left(P_{k+1} + \Delta P_{k+1}^{IT}\right) H_{k+1}^{IT^T} R_{k+1}^{-1} \Big\{ \nu_{k+1} + H_{k+1} \Big(\Phi_k \hat{x}_k^{IT} + G_k q_k \\
&+ \left[\mathcal{E}_k : x_k x_k^T - \hat{x}_k^{IT} \hat{x}_k^{IT}\right]\Big) + \left[E_{k+1} : x_{k+1} x_{k+1}^T - \hat{x}_{k+1}^{IT} \hat{x}_{k+1}^{IT^T}\right] \\
&+ 2\left[E_{k+1} : \hat{x}_{k+1}^{IT}\right] P_{k+1} H_{k+1}^T R_{k+1}^{-1} \Big(\nu_{k+1} + H_{k+1} \Big(\Phi_k \tilde{x}_k^{IT} + G_k q_k \\
&+ \left[\mathcal{E}_k : x_k x_k^T - \hat{x}_k^{IT} \hat{x}_k^{IT^T}\right]\Big) + \left[E_{k+1} : x_{k+1} x_{k+1}^T - \hat{x}_{k+1/k}^{IT} \hat{x}_{k+1/k}^{IT^T}\right]\Big)\Big\}
\end{aligned} \tag{3.40}$$

The term ΔP_{k+1}^{IT} can be computed from the same equations used to compute ΔP_{k+1}^{EK}, namely (3.32b-c), with the driving term Γ being replaced by

$$\begin{aligned}
\Gamma_{k+1} = &-2P_{k+1/k} H_{k+1}^T M_{k+1}^{-1} \left[E_{k+1} : \hat{x}_{k+1/k+1}\right] P_{k+1} \\
&-2P_{k+1} \left[E_{k+1} : \hat{x}_{k+1/k+1}\right]^T M_{k+1}^{-1} H_{k+1} P_{k+1/k}
\end{aligned} \tag{3.41}$$

As before, (3.40) is solved by replacing the small terms with their linear equivalents and using the properties of these random variables to obtain a forward difference equation for the iterated filter bias:

$$E\left\{\tilde{x}^{IT}_{k+1}\right\} = \tilde{\Phi}_{k+1/k}\, E\left\{\tilde{x}^{IT}_{k}\right\} + \left(I - P_{k+1} H^T_{k+1} R^{-1}_{k+1} H_{k+1}\right)\left[\mathcal{E}_k : P_k\right]$$

$$- P_{k+1}\left\{2\left[E^T_{k+1} : \left(P_{k+1} H^T_{k+1} R^{-1}_{k+1}\right)\right] + H^T_{k+1} R^{-1}_{k+1}\left[E_{k+1} : P_{k+1}\right]\right\}$$

$$E\left\{\tilde{x}^{IT}_{0}\right\} = 0 \qquad (3.42)$$

The covariance of this filter will again be the same as that of the other three algorithms.

PRECEDING PAGE BLANK NOT FILMED

4. EXAMPLES

Introduction

In order to gain some insight into the use of the bias expressions given in the last chapter, some simple examples are presented. The first deals with the general problem of parameter estimation and compares the results obtained by Breakwell (Ref. 5) with those in Chapter 3. In the second example the bias associated with applying the least square, extended Kalman, or iterated algorithm to a simple one dimensional parameter estimation scheme is carried out numerically.

The third and fourth examples consider a simple dynamic plant, with observation and dynamic non-linearities, respectively. These bias effects are considered separately since they simply superpose when both are present. It should be emphasized that the biases associated with any particular problem are very much a function of such parameters as initial covariances. Hence, it is dangerous to draw generalizations from the numerical data presented here for other problems.

4.1 Bias in Parameter Estimation with Slight Non-Linearity

The simplest type of system for which the equations in Chapter 3 are relevant consist of a set of constants which are elements of a vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \tag{4.1}$$

whose values are to be estimated from the (slightly non-linear) observations

$$y_k = H_k x_k + \left[E_k : x_k x_k^T\right] + \nu_k \qquad k = 1, 2, \ldots, N \tag{4.2}$$

where E_k is small. In this case Eqs. (3.15), (3.35) and (3.42) which give the estimation biases for the LS, EK or IT algorithms simplify considerably. In fact the transition matrix which occurs in the error transition matrix, defined in (3.16), becomes the identity matrix. Furthermore, only one index in the covariance is necessary so

$$P_{\ell/N} \triangleq P_N \tag{4.3}$$

With these relationships we further simplify by writing

$$\left(I - P_{\ell+1} H_{\ell+1}^T R_{\ell+1}^{-1} H_{\ell+1}\right) = P_{\ell+1}\left(P_{\ell+1}^{-1} - H_{\ell+1} R_{\ell+1}^{-1} H_{\ell+1}\right) = P_{\ell+1} P_{\ell}^{-1} \tag{4.4}$$

When these relations are used in the bias equations mentioned above with each equation multiplied through by $P_{\ell+1}^{-1}$, they become

$$P_{\ell+1}^{-1} E\left\{\tilde{x}_{\ell+1}^{LS}\right\} = P_{\ell}^{-1} E\left\{\tilde{x}_{\ell}^{LS}\right\} - 2\left[E_{\ell+1}^T : \left(P_N H_{\ell+1}^T R_{\ell+1}^{-1}\right)\right] - H_{\ell+1}^T R_{\ell+1}^{-1}\left[E_{\ell+1} : P_N\right] \tag{4.5a}$$

$$P_{\ell+1}^{-1} E\left\{\tilde{x}_{\ell+1}^{EK}\right\} = P_{\ell}^{-1} E\left\{\tilde{x}_{\ell}^{EK}\right\} - H_{\ell+1}^T R_{\ell+1}^{-1}\left[E_{\ell+1} : P_{\ell}\right] \tag{4.5b}$$

$$\begin{aligned} P_{\ell+1}^{-1} E\left\{\tilde{x}_{\ell+1}^{IT}\right\} &= P_{\ell}^{-1} E\left\{\tilde{x}_{\ell}^{IT}\right\} - 2\left[E_{\ell+1}^T : \left(P_{\ell+1} H_{\ell+1}^T R_{\ell+1}^{-1}\right)\right] \\ &\quad - H_{\ell+1}^T R_{\ell+1}^{-1}\left[E_{\ell+1} : P_{\ell+1}\right] \end{aligned} \tag{4.5c}$$

with zero initial conditions on the biases. To compare this with the results obtained by Breakwell (Ref. 5), the equations are solved and written as

$$E\left\{\tilde{x}_N^{LS}\right\} = -P_N \sum_{j=1}^{N} \left(H_j^T R_j^{-1}\left[E_j : P_N\right] + 2\left[E_j^T : \left(P_N H_j^T R_j^{-1}\right)\right]\right) \tag{4.6a}$$

$$E\left\{\tilde{x}_N^{EK}\right\} = -P_N \sum_{j=1}^{N} \left(H_j^T R_j^{-1} \left[E_j : P_{j-1}\right]\right) \tag{4.6b}$$

$$E\left\{\tilde{x}_N^{IT}\right\} = -P_N \sum_{j=1}^{N} \left(H_j^T R_j^{-1} \left[E_j : P_j\right] + 2\left[E_j^T : \left(P_j H_j^T R_j^{-1}\right)\right]\right) \tag{4.6c}$$

Now Breakwell has derived bias equations for the case where the y_j are scalars (so that R_j is a scalar) and R_j is unity. It turns out that Eqs. (4.6a-c) do not bear a one to one correspondence with Breakwell's. The difference, however, is entirely attributable to differences in assumptions about the meaning of the a priori covariance P. In the development of Eqs. (4.6a,b,c) it was assumed that the true initial state x_0 was a sample drawn from a group having a probability distribution with a known mean $\bar{x}_0$ and covariance P_0. This may or may not be a good mathematical model for a particular situation. It does enjoy the advantage, however, of giving the information necessary to use a Bayesian approach to utilizing a priori information to start the estimation schemes.

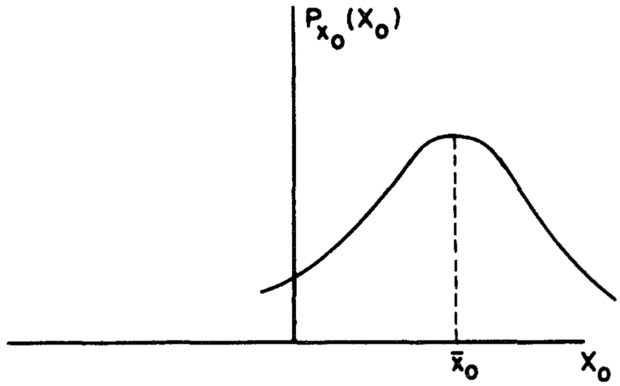
On the other hand, Breakwell started with the interesting assumption that $\bar{x}_0$ is a "pseudo-measurement" of x_0 contaminated by zero mean noise having covariance P_0. Thus the error $\tilde{x}_0$ is interpreted as being centered around the unknown mean x_0, whereas in the Bayesian approach it is centered around the known mean $\bar{x}_0$. Figure 1 schematizes the assumptions made in each of the approaches.

An alternate way of looking at the difference between these two approaches is to say that Breakwell computes the bias conditioned on knowing what the true value of the a priori state is. That is, he computes the condittional mean

$$E\left\{x_k - \hat{x}_k / x_0\right\}$$

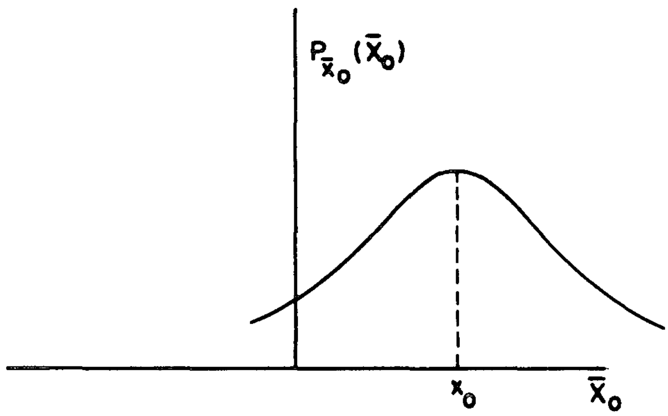
whereas the means computed in Chapter 3 are based on

$$E\left\{x_k - \hat{x}_k / E\left\{x_0\right\} = \bar{x}_0\right\}$$



(a)

Bayesian: $\bar{x}_0$ known and interpreted as mean of random variable x_0



(b)

Pseudo-measurement: a sample $\bar{x}_0$ is known and is assumed distributed with zero mean about true x_0

Fig. 1. DIFFERENCE BETWEEN PSEUDO-MEASUREMENT AND BAYESIAN APPROACH TO A PRIORI KNOWLEDGE.

When Breakwell's results are generalized and separated to obtain a form comparable to Eqs. (4.6a, b, and c) they look the same except for an additional additive term given by

$$2P_N \sum_{j=1}^{N} \left(H_j^T R_j^{-1} \left[E_j : P_N \right] + \left[E_j^T : \left(P_N H_j R_j^{-1} \right) \right] \right) \tag{4.7}$$

This term, when added to each of Eqs. (4.6a,b,c) does not change the basic conclusions reached regarding the long term behavior of the biases but it does change the sign and possible conclusions regarding the bias in the case of little data.

With reference to Eqs. (4.6a,b,c) we can make many of the comments made by Breakwell. For example, although the least square weights all the non-linearities with P_N, and the iterated weights them by P_j (both of which are presumably smaller than P_{j-1} which occurs in the extended Kalman filter) there are three times as many terms in the former cases, so that the last may actually be the best. A simple example which points this out is investigated in the next section.

4.2 A Numerical Example of Scalar Parameter Estimation Bias with Small Non-Linearity

The simplest example which can be considered is the problem of estimating a scalar constant c using observations

$$y_i = c + \epsilon c^2 + \nu_i \qquad i = 1, 2, \ldots, N \tag{4.8}$$

where ν_i is a zero mean white noise process with covariance r and c has a priori covariance p_0. The inverse covariance update equation gives

$$p_k = \frac{1}{\frac{1}{p_0} + \frac{k}{r}} \tag{4.9}$$

Hence, making the proper substitutions in Eqs. (4.6a, b, and c), all of whose terms are scalars for this example, the biases become simply:

$$E\left\{\tilde{c}_N^{LS}\right\} = -\frac{3r}{\left(\frac{r}{p_0} + N\right)^2} \sum_{j=1}^{N} \epsilon \triangleq C_N \cdot 3 \sum_{j=1}^{N} \frac{1}{\alpha + N} \tag{4.10a}$$

$$E\left\{\tilde{c}_N^{EK}\right\} = -\frac{r}{\left(\frac{r}{p_0} + N\right)} \sum_{j=1}^{N} \frac{\epsilon}{\left(\frac{r}{p_0} + j - 1\right)} \triangleq C_N \sum_{j=1}^{N} \frac{1}{\alpha + j - 1} \tag{4.10b}$$

$$E\left\{\tilde{c}_N^{IT}\right\} = -\frac{3r}{\left(\frac{r}{p_0} + N\right)} \sum_{j=1}^{N} \frac{\epsilon}{\left(\frac{r}{p_0} + j\right)} \triangleq C_N \cdot 3 \sum_{j=1}^{N} \frac{1}{\alpha + j} \tag{4.10c}$$

The relative sizes of the biases have now been reduced to a function of two parameters, the ratio r/p_0 defined here to be α, and the number of observations N. The LS and IT bias have an unfavorable factor of three (3) as compared to the EK filter, but the individual denominator terms in the EK filter are larger, reflecting the fact that a predicted covariance is used in each term.

For large N, i.e., $N \gg \alpha$, the term which gives the relative bias of the LS algorithm is approximated by

$$3 \sum_{j=1}^{N} \frac{1}{\alpha + N} = \frac{3N}{\alpha + N} \cong 3 \tag{4.11}$$

The pertinent term in the iterated filter can be evaluated by noting that the summation looks much like a coarse partition used to evaluate the Riemann integral

$$\int_1^{N+1} \frac{1}{\alpha + x}\, dx$$

Simple geometric considerations then lead us to write

$$\ell n\ (N + 1 + \alpha) + \ell n\ (1 + \alpha) \leq \sum_{j=1}^{N} \frac{1}{\alpha + j} \leq \ell n\ (N + \alpha) - \ell n\ (1 + \alpha) + \frac{1}{\alpha + 1}$$

Hence, for

$$N \gg 1 + \alpha$$

$$3 \sum_{j=1}^{N} \frac{1}{\alpha + j} \cong 3\ \ell n\ (N) \tag{4.12}$$

The EK summation term can be similarly evaluated by writing

$$\sum_{j=1}^{N} \frac{1}{\alpha + j - 1} = \frac{1}{\alpha} - \frac{1}{\alpha + N} + \sum_{j=1}^{N} \frac{1}{\alpha + j} \sim \frac{1}{\alpha} + \ell n\ (N) \tag{4.13}$$

$$N \gg 1 + \alpha$$

From Eqs. (4.11-4.13) we see that for large enough N, the LS bias will be the smallest. When $\ell n\ N \gg 1/\alpha$, the EK bias will be about one-third of the iterated bias so that iteration would probably not be as advantageous as using the simpler EK. However, the term $1/\alpha$ is likely to be large, reflecting a large a priori uncertainty relative to the weighting put on data. For example, if p_0 is taken as ten times as large as r, $1/\alpha$ takes the value ten (10). Hence, N must take on the value 10,000 before $\ell n\ (N)$ even reaches a magnitude comparable to $1/\alpha$. For any particular problem the point where the EK bias becomes less than the iterated bias is likely to be more interesting than the limiting case. For small α this will occur when

$$\ell n\ N \cong \frac{1}{2\alpha} \tag{4.14}$$

For moderate values of N and α, it is entirely possible that the coefficient 3 in Eqs. (4.10a-c) is a dominating influence so that the simple EK algorithm would have the lowest biases.

Figure 2* depicts the bias histories for $E = .1$, an initial covariance P_0 chosen as .1, and three noise variance cases, with respective values of .009, .01, .011. This gives fairly small values of α (.09, .1 and .11 respectively). In these cases the least square bias is always the smallest. The iterative bias is less than the extended Kalman initially, but after a while, the EK filter actually has the smaller bias. The crossover actually occurs at about 170, 90 and 65 points, respectively whereas the rough approximation given by (4.14) would have placed the crossovers at 245, 150 and 90 points.

The interesting case where the EK filter is actually the least biased initially is plotted in Fig. 3. The case depicted here represents a measurement noise covariance of .1. That is, the *a priori* covariance contains about the same information as a data point so that $\alpha = 1.0$. For this α, the iterated filter always has a larger bias than the EK. The least square estimator always becomes the least biased after a sufficient number of data points are considered.

4.3 A Dynamic Example with Small Observation Non-Linearity

A simple two state dynamic system with small observation non-linearity can be written in terms of the two scalar difference equations

$$z_{k+1} = \varphi z_k + (1 - \varphi)\, c_k + q_k$$
$$c_{k+1} = c_k \qquad (4.15a)$$

*Since logarithmic scales were most convenient, the negative value of the biases is plotted in Fig. 2.

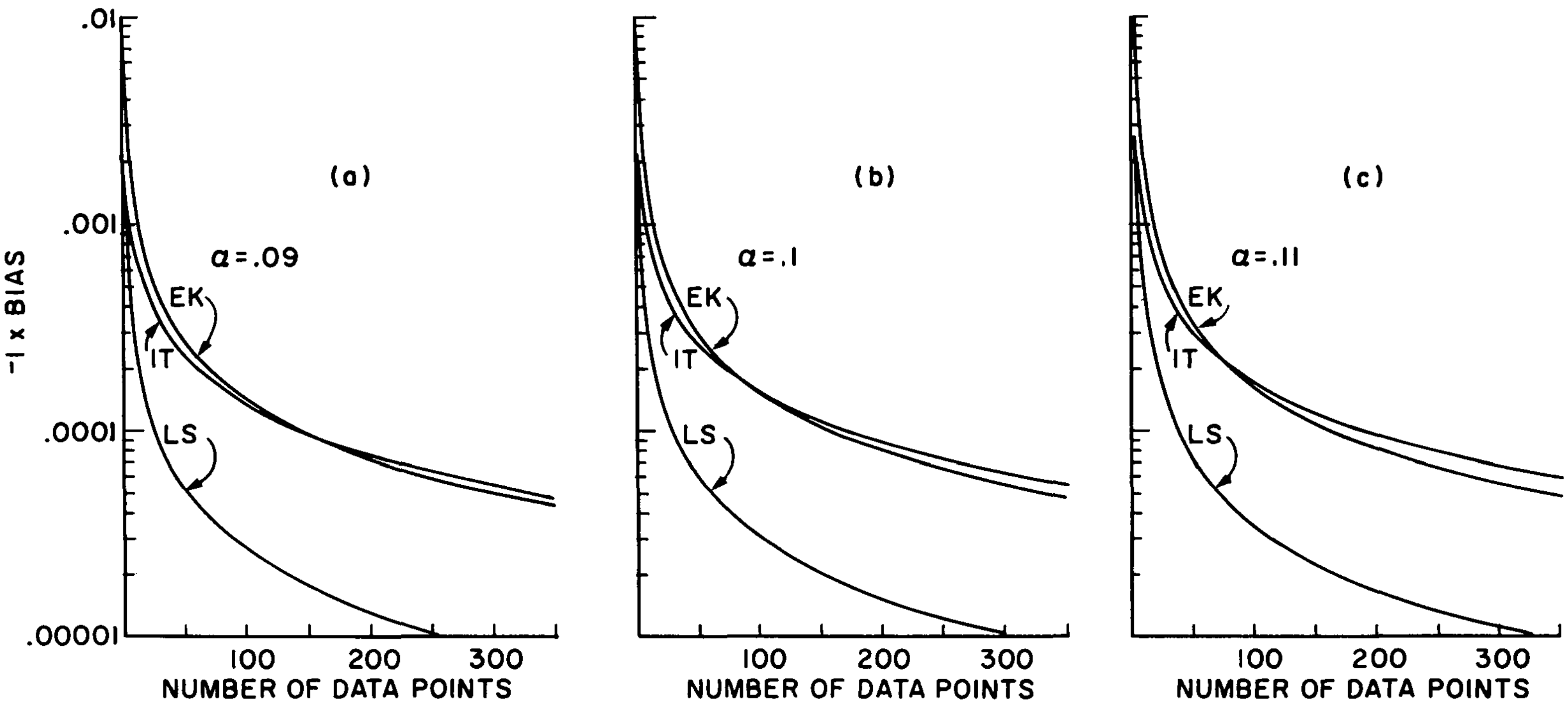


Fig. 2. BIAS HISTORIES FOR STATIC CASE WITH SMALL ALPHA.

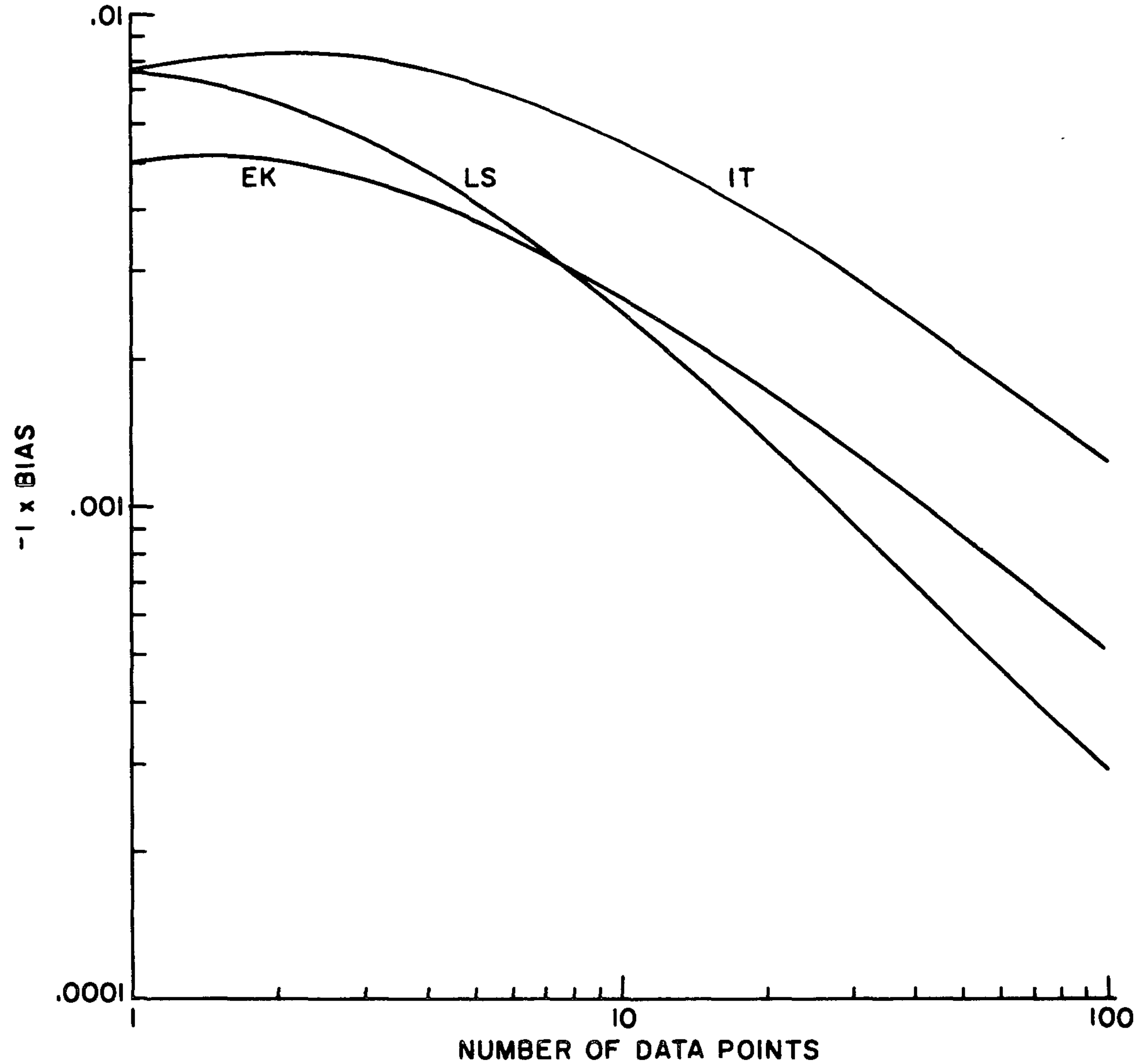


Fig. 3. NEGATIVE OF BIAS HISTORIES FOR STATIC EXAMPLE WITH LARGE α (= 1.0).

and an observation equation

$$y_k = z_k + \epsilon z_k^2 + \nu_k \tag{4.15b}$$

Such a mathematical model might be used to describe a physical system designed to measure the constant c where the transducer is a first order plant with unity steady state gain. The white, zero mean noise q_k (having covariance Q) is a perturbation at the input of the transducer, whereas the ν_k is a white noise with covariance R corrupting the output. If we ignored the non-linearity ϵ and transient effects, this would be a problem in extracting the mean c from a series of data points containing both colored noise due to q_k and white noise due to ν_k.

The covariances associated with processing the data y_k when $\epsilon = 0$ (the linear problem) are independent of the data y_k and will be defined by

$$E\left\{\tilde{z}_{k/N}\tilde{z}_{k/N}\right\} \triangleq p_{zz}(k/N) \tag{4.16a}$$

$$E\left\{\tilde{c}_N\tilde{c}_N\right\} \triangleq p_{cc}(N) \tag{4.16b}$$

$$E\left\{\tilde{z}_{k/N}\tilde{c}_N\right\} \triangleq p_{cz}(k/N) \tag{4.16c}$$

The index k has been dropped in (4.16b) since c is a constant. One index will also be dropped when both take on the same value. With these definitions the bias Eqs. (3.15), (3.35) and (3.42) reduce to the three pairs of coupled difference equations summarized in Table 1. These computations were carried out numerically for the case where $\varphi = 0.95$, $\epsilon = 0.1$, $p_{zz}(0)^*$ = 0.20, $p_{cz}(0) = 0.0$, $p_{cc}(0) = .02$ and R was

---

*Some cases were also run with $p_{cc}(0) = 0.2$. These gave results very similar to those obtained using the smaller value.

Table 1

BIAS EQUATIONS FOR SMALL OBSERVATION NON-LINEARITIES

Least Square Bias

$$E\left\{\tilde{z}_{k+1}^{LS}\right\} = \tilde{\varphi}_{zz}(k)\ E\left\{\tilde{z}_{k}^{LS}\right\} + \tilde{\varphi}_{zc}(k)\ E\left\{\tilde{c}_{k}^{LS}\right\} - \frac{3\epsilon}{R}\ p_{zz}(k+1)p_{zz}(k+1/N)$$

$$E\left\{\tilde{c}_{k+1}^{LS}\right\} = \tilde{\varphi}_{cz}(k)\ E\left\{\tilde{z}_{k}^{LS}\right\} + \tilde{\varphi}_{cc}(k)\ E\left\{\tilde{c}_{k}^{LS}\right\} - \frac{3\epsilon}{R}\ p_{zc}(k+1)p_{zz}(k+1/N)$$

Extended Kalman Bias

$$E\left\{\tilde{z}_{k+1}^{EK}\right\} = \tilde{\varphi}_{zz}(k)\ E\left\{\tilde{z}_{k}^{EK}\right\} + \tilde{\varphi}_{zc}(k)\ E\left\{\tilde{c}_{k}^{EK}\right\} - \frac{\epsilon}{R}\ p_{zz}(k+1)p_{zz}(k+1/k)$$

$$E\left\{\tilde{z}_{k+1}^{EK}\right\} = \tilde{\varphi}_{cz}(k)\ E\left\{\tilde{z}_{k}^{EK}\right\} + \tilde{\varphi}_{cc}(k)\ E\left\{\tilde{c}_{k}^{LS}\right\} - \frac{\epsilon}{R}\ p_{zc}(k+1)p_{zz}(k+1/k)$$

Iterated Bias

$$E\left\{\tilde{z}_{k+1}^{IT}\right\} = \tilde{\varphi}_{zz}(k)\ E\left\{\tilde{z}_{k}^{IT}\right\} + \tilde{\varphi}_{zc}(k)\ E\left\{\tilde{c}_{c}^{IT}\right\} - \frac{3\epsilon}{R}\ p_{zz}(k+1)p_{zz}(k+1)$$

$$E\left\{\tilde{c}_{k+1}^{IT}\right\} = \tilde{\varphi}_{cz}(k)\ E\left\{\tilde{z}_{k}^{IT}\right\} + \tilde{\varphi}_{cc}(k)\ E\left\{\tilde{c}_{c}^{IT}\right\} - \frac{3\epsilon}{R}\ p_{zc}(k+1)p_{zz}(k+1)$$

where

$$\tilde{\varphi}_{zz}(k) \triangleq \left(1 - \frac{p_{zz}(k+1)}{R}\right)\varphi \qquad \tilde{\varphi}_{zc}(k) \triangleq \left(1 - \frac{p_{zz}(k+1)}{R}\right)(1-\varphi)$$

$$\tilde{\varphi}_{cz}(k) \triangleq -\frac{p_{cz}(k+1)}{R}\varphi \qquad \tilde{\varphi}_{cc}(k) = 1 - \frac{p_{cz}(k+1)}{R} \quad (1-\varphi)$$

$$E\left\{\tilde{z}_{0}^{LS}\right\} = E\left\{\tilde{z}_{0}^{EK}\right\} = E\left\{\tilde{z}_{0}^{IT}\right\} = 0$$

$$E\left\{\tilde{c}_{0}^{LS}\right\} = E\left\{\tilde{c}_{0}^{EK}\right\} = E\left\{\tilde{c}_{0}^{IT}\right\} = 0$$

chosen as .04. Figure 4 depicts the bias in the estimation of c for three cases: no state noise $(Q = 0)$, $Q = .0001$ and $Q = .001$. With only a few data points, the least square estimator is always the least biased, the iterated is second, and the extended Kalman, the worst. This is attributable to the fact that the *a priori* covariance is weighted out the fastest by the LS estimator, and the slowest by the EK estimator. However, as the dependence on initial covariances dies out, the bias seems to go through a region where the factor of three in the IT and LS algorithms makes their bias larger than that of the EK. This is in contradistinction to the static example where the LS estimate always has the smallest bias.

For the case where $Q = 0$, depicted in 4a, some special calculations were carried out to confirm the fact that eventually the LS bias would become smaller than the EK. This happened at between 11000 and 12000 data points.

The effect of state noise seems twofold. Initially it reduces the bias over the non-state noise case by damping out the effect of *a priori* information. When a large number of data points are used however, it has the opposite effect, since all the non-linearities become weighted by larger values reflecting the fact that the state noise forces increased estimate uncertainties. It appears that the region where the EK estimate is better than the LS estimate is likely to be much larger for increasing state noise covariance.

4.4 A Dynamic Example with Small Dynamic Non-Linearity

This section, like the last, considers a simple two state model, but with a dynamic rather than observational non-linearity. The dynamic equations are

$$z_{k+1} = \varphi z_k + \epsilon z_k^2 + (1 - \varphi)\, c_k + q_k$$

$$c_{k+1} = c_k \tag{4.17a}$$

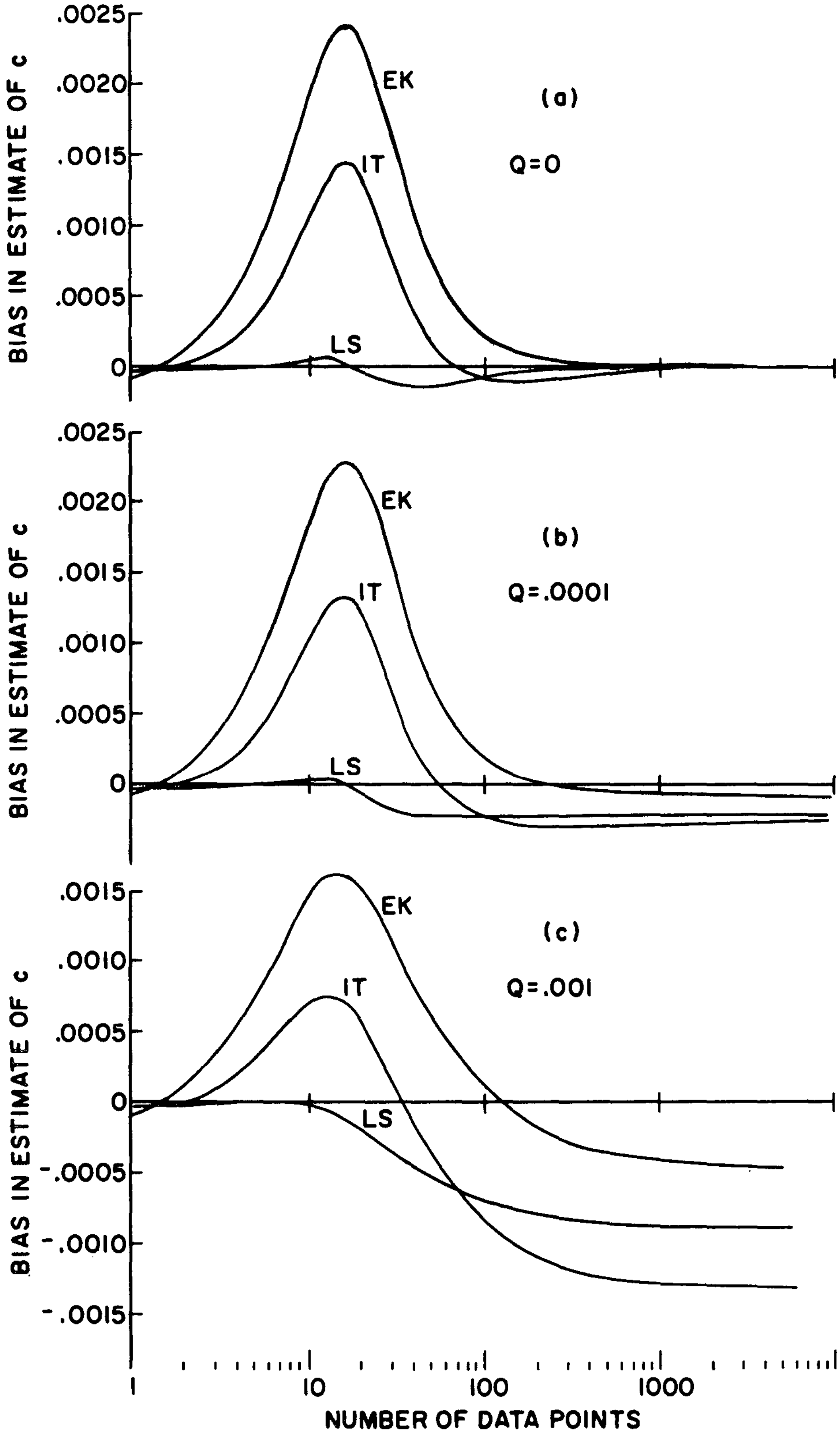


Fig. 4. BIAS HISTORIES FOR OBSERVATIONAL NON-LINEARITY IN DYNAMIC PROBLEM.

The observation equation is simply

$$y_k = z_k + \nu_k \tag{4.17b}$$

The relevant bias equations are summarized in Table 2. The parameters are identical with those given in Section 4.3. ϵ is chosen as .1. In the case of dynamic non-linearities, the EK and IT biases are identical. In Fig. 5 these are compared with the LS bias, which, for any given value of state noise, is always the smallest.

Table 2

BIAS EQUATIONS FOR SMALL DYNAMIC NON-LINEARITIES

Least Square Bias

$$E\left\{\tilde{z}^{LS}_{k+1}\right\} = \tilde{\varphi}_{zz}(k)\ E\left\{z^{LS}_{k}\right\} + \tilde{\varphi}_{zc}(k)\ E\left\{\tilde{c}^{LS}_{k}\right\} + \epsilon\left(1 - \frac{p_{zz}(k+1)}{R}\right) p_{zz}(k/N)$$

$$E\left\{\tilde{c}^{LS}_{k+1}\right\} = \tilde{\varphi}_{cz}(k)\ E\left\{\tilde{z}^{LS}_{k}\right\} + \tilde{\varphi}_{cc}(k)\ E\left\{\tilde{c}^{LS}_{k}\right\} - \epsilon\ p_{cz}(k+1)\ p_{zz}(k/N)$$

Extended Kalman and Iterated Bias

$$E\left\{\tilde{z}^{IT}_{k+1}\right\} = \tilde{\varphi}_{zz}(k)\ E\left\{\tilde{z}^{IT}_{k}\right\} + \tilde{\varphi}_{zc}(k)\ E\left\{\tilde{c}^{IT}_{k}\right\} + \epsilon\left(1 - \frac{p_{zz}(k+1)}{R}\right) p_{zz}(k)$$

$$E\left\{\tilde{z}^{IT}_{k+1}\right\} = \tilde{\varphi}_{cz}(k)\ E\left\{\tilde{z}^{IT}_{k}\right\} + \tilde{\varphi}_{cc}(k)\ E\left\{\tilde{c}^{IT}_{R}\right\} - \epsilon\ \frac{p_{cz}(k+1)}{R}\ p_{zz}(k)$$

$$\tilde{\varphi}_{zz}(k) \triangleq \left(1 - \frac{p_{zz}(k+1)}{R}\right)\varphi \qquad \tilde{\varphi}_{zc}(k) \triangleq \left(1 - \frac{p_{zz}(k+1)}{R}\right)(1 - \varphi)$$

$$\tilde{\varphi}_{cz}(k) \triangleq -\frac{p_{cz}(k+1)}{R}\varphi \qquad \tilde{\varphi}_{cc}(k) \triangleq 1 - \frac{p_{cz}(k+1)}{R}(1 - \varphi)$$

$$E\left\{\tilde{z}^{LS}_{0}\right\} = E\left\{\tilde{z}^{IT}_{0}\right\} = 0$$

$$E\left\{\tilde{c}^{LS}_{0}\right\} = E\left\{\tilde{c}^{IT}_{0}\right\} = 0$$

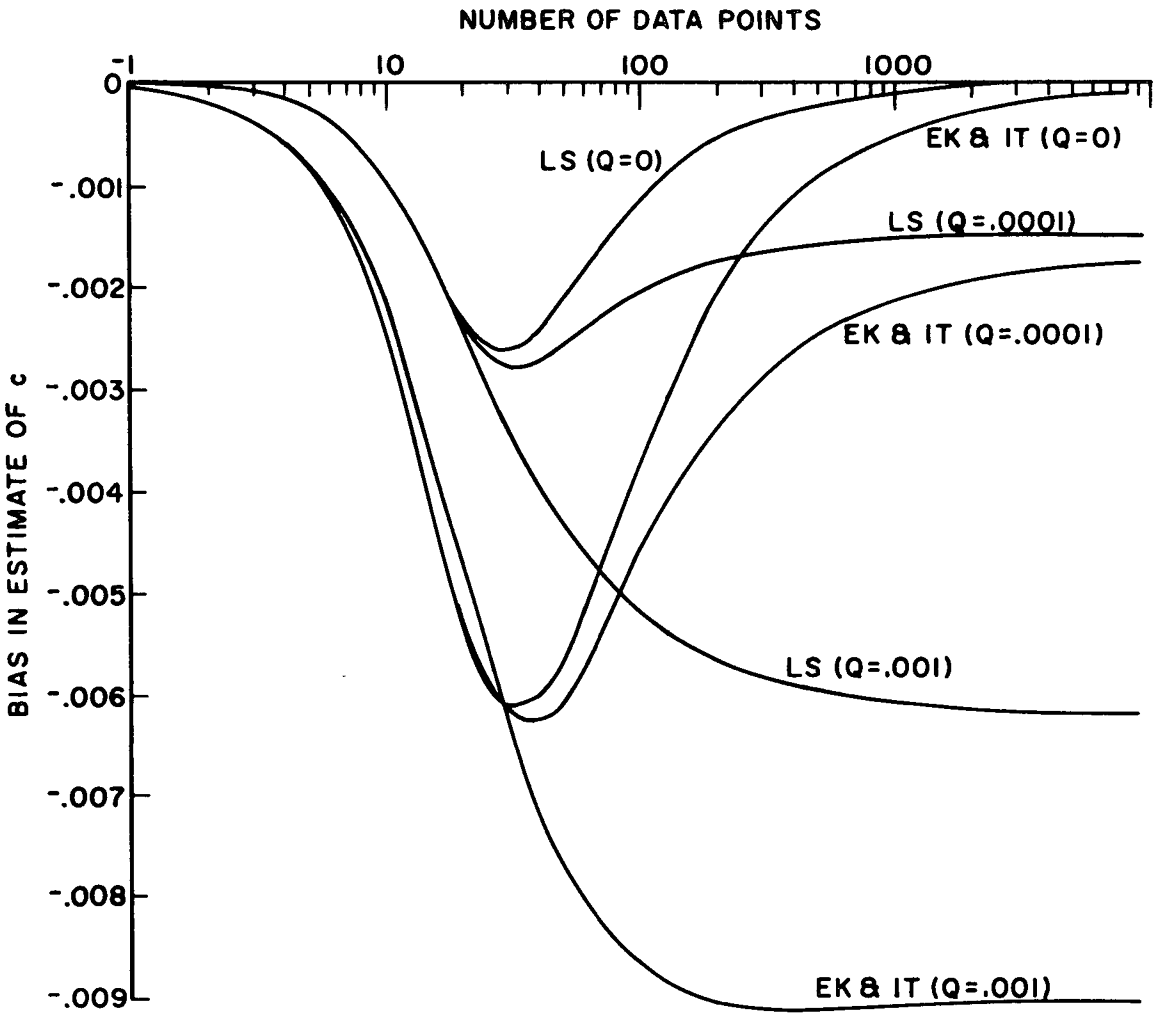


Fig. 5. BIAS HISTORIES FOR DYNAMIC NON-LINEARITIES.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The examples in Chapter 4 point out some interesting characteristics regarding estimation biases associated with using an extended Kalman (EK), iterated Kalman (IT), or least square (LS) algorithm to filter data from slightly non-linear dynamic systems. Several of these characteristics are in good agreement with results obtained by Breakwell for the parameter estimation problem. For example, for small amounts of data, the LS algorithm has the smallest bias, the EK filter has the largest bias, while the IT filter bias has an intermediate value. Another phenomenon which parallels Breakwell's observations is the tendency of the IT filter to give the largest bias after a large number of data points have been processed, even though state noise may be present. Very large values of state noise may give different results.

The dynamic problem possesses the interesting property that the LS bias due to observational nonlinearities can actually be larger than the EK bias after a large number of data points. With no state noise the LS will eventually be the least biased but this may require a vast number of data points. Numerical results indicate that, with state noise, the LS bias may never become as small as the EK bias.

In general, the bias equations derived in Chapter 5 show that for small dynamic non-linearities (as opposed to observational non-linearities) the EK and IT filters have the same bias. This bias will generally be larger than the LS bias.

Finally, a filter is developed that removes the biases inherent in the other algorithms. By carrying second order terms, both the effects of small dynamic and observational non-linearities on the estimate bias are removed.

5.2 Recommendations

The hypothesis required to obtain the convenient computational expressions derived in Chapter 3, involved fairly strong assumptions about the size of the variances of the random variables involved in the problem.

The net effect of these assumptions is to limit the size of the biases of the state estimates to a fraction of the covariance of the state estimates. This leaves an important question to be pursued.

The question concerns realistic problems with blatantly moderate size non-linearities. Although it is necessary to assume small non-linearities to get the fairly tractable results of Chapter 3, one hopes that relative characteristics of the different filters displayed in the examples in Chapter 4 carry over larger non-linearities. It would be interesting to take a particular problem and determine if this is true. This could be done using Monte-Carlo techniques if no analytic solution presented itself.

A second question concerns the impact of coordinate transformations on the algorithms. It would be interesting to determine the possible benefits of using coordinates where the non-linearities become either all dynamic or all observational, since the two have different effects. This may lead to a better iterative algorithm.

APPENDIX A

This section shows that the linear two point boundary value in Chapter 3 is the solution to a least-square problem associated with linear systems, and then derives the recursive solutions used in finding approximations to the slightly non-linear problem.

The equations are obtained by minimizing the quadratic performance criterion

$$\varphi = \frac{1}{2}\left[\bar{x}_0 - \hat{x}(0/N)\right]^T P_0^{-1}\left[\bar{x}_0 - \hat{x}(0/N)\right]$$

$$+ \frac{1}{2}\sum_{i=j}^{N}\left[y_i - H_i\hat{x}(i/N)\right]^T R_i^{-1}\left[y_i - H_i\hat{x}(i/N)\right] \tag{A.1}$$

$$+ \frac{1}{2}\sum_{i=0}^{N-1}\hat{q}_i Q_i^{-1}\hat{q}_i$$

with respect to the $\hat{x}$ and $\hat{q}$ subject to the constraint

$$\hat{x}(i + 1/N) = \Phi_i\hat{x}(i/N) + G_i\hat{q}_i + u_i \tag{A.2}$$

This problem is solved by adjoining the constraints (A.2) to the performance criteria with a sequence of Lagrange multipliers $\lambda(i/N)$ to obtain

$$\bar{\varphi} = \varphi + \sum_{i=0}^{N-1}\lambda^T(i/N)\left[\Phi_i\hat{x}(i/N) + G_i\hat{q}_i + u_i - \hat{x}(i + 1/N)\right] \tag{A.3}$$

With a little rearrangement, the last equation can be rewritten as

$$\bar{\varphi} = \frac{1}{2}\left[\bar{x}_0 - \hat{x}(0/N)\right]^T P_0^{-1}\left[\bar{x}_0 - \hat{x}(0/N)\right]$$

$$+ \sum_{i=1}^{N-1}\left\{\frac{1}{2}\left[y_i - H_i\hat{x}(i/N)\right]^T R_i^{-1}\left[y_i - H_i\hat{x}(i/N)\right] + \frac{1}{2}\hat{q}_i^T Q_i^{-1}\hat{q}_i\right.$$

$$\left. + \lambda^T(i/N)\left[\Phi_i\hat{x}(i/N) + G_i\hat{q}_i + u_i - \hat{x}(i+1/N)\right]\right\} \tag{A.4}$$

$$+ \lambda^T(0/N)\left[\Phi_0\hat{x}(0/N) + G_0\hat{q}_0\right] + \frac{1}{2}\left[y_N - H_N\hat{x}(N/N)\right]^T R_N^{-1}\left[y_N - H_N\hat{x}(N/N)\right]$$

$$+ \frac{1}{2}\hat{q}_0^T Q_0^{-1}\hat{q}_0 - \lambda_{N-1}^T\hat{x}(N/N)$$

Differentiating with respect to $\hat{q}_j$ and setting the result to zero gives

$$\hat{q}_j = -Q_j G_j^T \lambda(j/N) \qquad j = 0,1,--,N-1 \quad . \tag{A.5}$$

Differentiating with respect to $\hat{x}(j/N)$ $(j = 1,--,N-1)$

$$\lambda(j-1/N) = \Phi_j^T\lambda_j - H_j^T R_j^{-1}\left[y_j - H_j\hat{x}(j/N)\right] \quad . \tag{A.6}$$

Differentiating with respect to $\hat{x}(N/N)$ gives

$$\lambda(N-1/N) = -H_N^T R_N^{-1}\left[y_N - H_N\hat{x}(N/N)\right] \tag{A.7}$$

But equivalently, we can assume Eq. (A.6) holds for $(j = 1,--N)$ with $\lambda(N/N)$ taken as zero. Differentiation with respect to $\hat{x}(0/N)$ gives a final necessary relation:

$$-P_0^{-1}\left[\bar{x}_0 - \hat{x}(0/N)\right] + \Phi_0^T\lambda(0/N) = 0 \tag{A.8}$$

All the equations for solving the least square problem can now be summarized. Equation (A.5) is substituted into (A.2) to get one N^{th} order difference equation while (A.6) gives another. Equations (A.7) (with the accompanying remark) and (A.8) give 2N boundary conditions. In particular, we must solve

$$\hat{x}(i + 1/N) = \Phi_i\hat{x}(i/N) + u_i - G_iQ_iG_i^T\lambda(i/N) \tag{A.9}$$

$$\lambda(i - 1/N) = \Phi_i^T\lambda(i/N) - H_iR_i^{-1}\left[y_i - H_i\hat{x}(i/N)\right] \qquad i = 1, 2, --, N \tag{A.10}$$

with

$$\lambda(N/N) = 0 \tag{A.11}$$

$$P_0\Phi_0^T\lambda(0/N) = \bar{x}_0 - \hat{x}(0/N) \tag{A.12}$$

The recursive solutions to Eqs. (A.9) through (A.12) can be found using sweep methods suggested by Bryson (Ref. 7). This is done by breaking the solutions of (A.9) and (A.10) into homogeneous $(y_i = 0)$ and particular solutions. To do this, we first rewrite Eq. (A.10) in forward form:

$$\begin{aligned}\lambda(i + 1/N) = {} & \Phi_{i+1}^{-T}\left(I + H_{i+1}^TR_{i+1}^{-1}H_{i+1}G_iQ_iG_i^T\right)\lambda(i/N) \\ & + \Phi_{i+1}^{-T}H_{i+1}^TR_{i+1}^{-1}\left\{y_{i+1} - H_{i+1}\left[\Phi_i\hat{x}(i/N) + u_i\right]\right\}\end{aligned} \tag{A.13}$$

and consider a set of particular solutions $\hat{x}_p(i)$ and $\lambda_p(i)$ satisfying (A.9) and (A.13) with initial conditions.

$$\hat{x}_p(0) = \bar{x}_0$$

$$\lambda_p(0) = 0 \quad . \tag{A.14}$$

The solutions are then written as the sum of the particular solutions and the homogeneous solutions $\hat{x}_h(i/N)$ and $\lambda_h(i/N)$ so that

$$\hat{x}(i/N) = \hat{x}_h(i/N) + \hat{x}_p(i)$$

$$\lambda(i/N) = \lambda_h(i/N) + \lambda_p(i) \tag{A.15}$$

When (A.14) is substituted into (A.15) and the result used in the boundary condition (A.12), the relation between the initial conditions on the homogeneous solutions becomes

$$P_0\Phi_0^T\lambda_h(0/N) = -\hat{x}_h(0/N) \tag{A.16}$$

Equation (A.16) shows that the whole $2n \times 2n$ transition matrix associated with (A.14) and (A.16) need not be computed to find all pertinent homogeneous solutions. Instead, we can consider only those columns corresponding to

$$\lambda_h = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \quad \lambda_h = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \; . \; . \; .$$

with

$$\hat{x}_h = -P_0\Phi_0^T \begin{bmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \quad \hat{x}_h = -P_0\Phi_0^T \begin{bmatrix} 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}, \quad \ldots$$

so we need only solve for the reduced $2n \times n$ transition matrix solving (A.14) and (A.16) which is defined by

$$\begin{bmatrix} X(k) \\ \Lambda(k) \end{bmatrix} \qquad k = 0, 1, \text{--}, N \tag{A.17}$$

and satisfies the initial condition:

$$\begin{bmatrix} X(0) \\ \Lambda(0) \end{bmatrix} = \begin{bmatrix} -P_0\Phi_0^T \\ I \end{bmatrix} . \tag{A.18}$$

With these relations the homogeneous solution can be written as

$$\begin{aligned} \hat{x}_h(k/N) &= X(k)\ \lambda_h(0/N) \\ \lambda_h(k/N) &= \Lambda(k)\ \lambda_h(0/N) \end{aligned} \tag{A.19}$$

In particular, the terminal values are

$$\begin{aligned} \hat{x}_h(N/N) &= X(N)\ \lambda_h(0/N) \\ \lambda_h(N/N) &= \Lambda(N)\ \lambda_h(0/N) \quad . \end{aligned} \tag{A.20}$$

The terminal condition (A.11) requires that

$$\lambda_h(N/N) + \lambda_p(N) = 0$$

or

$$\lambda_h(N/N) = -\lambda_p(N) \quad . \tag{A.21}$$

Relations (A.20) and (A.21) can be combined to get

$$\hat{x}_h(N/N) = -X(N)\, \Lambda^{-1}(N)\, \lambda_p(N) \tag{A.22}$$

so that the desired solution to the TPBVP is

$$\hat{x}(N/N) = \hat{x}_p(N) + \hat{x}_h(N/N) = \hat{x}_p(N) - X(N)\, \Lambda^{-1}(N)\, \lambda_p(N) \quad , \tag{A.23}$$

where all the terms in Eq. (A.23) are computed in a forward computational pass. We can note from the symmetry of this equation that

$$\hat{x}(j/j) \triangleq \hat{x}_p(j) - X(j)\, \Lambda^{-1}(j)\, \lambda_p(j) \tag{A.24}$$

is a forward formula which satisfies the TPBVP where summations involve only j pieces of data.

The well known recursive solutions for (A.23) can be obtained by defining

$$P_{j+1} = -X(j+1)\, \Lambda^{-1}(j+1)\, \Phi_{j+1}^{-T} \quad . \tag{A.25}$$

From the homogeneous versions of (A.9) and (A.13) we deduce

$$P_{j+1} = -\left[\Phi_j X(j) - G_j Q_j G_j^T \Lambda(j)\right]$$

$$\cdot \left\{\Phi_{j+1}^{-T}\left[\left(I + S_{j+1} G_j Q_j G_j^T\right)\Lambda(j) - S_{j+1}\Phi_j X(j)\right]\right\}^{-1} \Phi_{j+1}^{-T}$$

(A.26)

$$= \left[-\Phi_j X(j)\Lambda^{-1}(j) + G_j Q_j G_j^T\right] \cdot \left[\left(I + S_{j+1} G_j Q_j G_j^T\right) - S_{j+1}\Phi_j X(j)\Lambda^{-1}(j)\right]$$

$$= \left[\Phi_j P_j \Phi_j^T + G_j Q_j G_j^T\right]\left[I + S_{j+1}\left(\Phi_j P_j \Phi_j^T + G_j Q_j G_j^T\right)\right]^{-1}$$

For convenience we can introduce the term

$$P_{j+1/j} \triangleq \Phi_j P_j \Phi_j^T + G_j Q_j G_j^T \tag{A.27}$$

so that

$$P_{j+1} = P_{j+1/j}\left(I + S_{j+1} P_{j+1/j}\right)^{-1}$$

(A.28)

$$= P_{j+1/j} - P_{j+1/j} H_{j+1}^T \left(H_{j+1} P_{j+1/j} H_{j+1}^T + R_{j+1}\right)^{-1} H_{j+1} P_{j+1/j}$$

Equation (A.25) is used in (A.24) to obtain

$$\hat{x}(j + 1/j + 1) = \hat{x}_p(j + 1) + P_{j+1}\Phi_{j+1}^T \lambda_p(j + 1) \tag{A.29}$$

But in (A.29) we can substitute for the particular solutions from (A.9) and (A.13) to get

$$\hat{x}(j + 1/j + 1) = \Phi_j \hat{x}_p(j) + u_j - G_j Q_j G_j^T \lambda_p(j)$$

$$+ P_{j+1}\Phi_{j+1}^T \Phi_{j+1}^{-T}\left(I - S_{j+1} G_j Q_j G_j^T\right)\lambda_p(j) \tag{A.30}$$

$$+ P_{j+1}\Phi_{j+1}^T \Phi_{j+1}^{-T} H_{j+1}^T R_{j+1}^{-1}\left\{y_{j+1} - H_{j+1}\left[\Phi_j x_p(j) + u_j\right]\right\}$$

When Eqs. (A.27) and (A.28) are used in (A.30) and the result regrouped, it can be seen that

$$\begin{aligned}\hat{x}(j+1/j+1) &= \Phi_j\left[\hat{x}_p(j) + P_j\Phi_j^T\lambda_p(j)\right] + u_j \\ &\quad + P_{j+1}H_{j+1}^T R_{j+1}^{-1}\left\{y_{j+1} - H_{j+1}^{-1}\left(\Phi_j\left[\hat{x}_p(j) + P_j\Phi_j^T\lambda_p(j)\right] + u_j\right)\right\} \\ &= \Phi_j\hat{x}(j/j) + u_j + P_{j+1}H_{j+1}^T R_{j+1}^{-1}\left\{y_{j+1} - H_{j+1}\left[\Phi_j\hat{x}(j/j) + u_j\right]\right\}\end{aligned} \tag{A.31}$$

In the sequel we define

$$\hat{x}(j+1/j) = \Phi_j\hat{x}(j/j) + u_j \tag{A.32}$$

The forward solution of Eqs. (A.27), (A.28), (A.31) and (A.32) with initial conditions P_0 and

$$\hat{x}(0/0) = \bar{x}_0 \tag{A.33}$$

for N points gives a solution to the TPBVP for $\hat{x}(N/N)$. Using $\lambda_N = 0$ we would now simply solve (A.9) and (A.10) backwards to obtain $\hat{x}(j/N)$ and $\lambda(j/N)$, $j = N-1, N-2, --, 0$. However, these backwards solutions tend to be numerically unstable and are not in a convenient form for the purposes of Chapter 2. Hence, we shall derive several alternative formulas for finding the smoothed values $\hat{x}(j/N)$. These formulas will be functions of the filtered quantities $\hat{x}(j/j)$.

Fixed Point Forward Smoothing

We will now find a solution for $\hat{x}(k/N)$ in terms of $\hat{x}(k/N\text{-}1)$ and $\hat{x}(k+1/k)$. We begin by assuming solutions for $\hat{x}(k/N\text{-}1)$ which are obtained from using all the relations obtained above with the index N replaced by $N-1$. To get an expression for the N^{th} term Eq. (A.21) is substituted into (A.20) and the resulting equation is solved for $\lambda_h(0/N)$. This is substituted into (A.19):

$$\hat{x}_h(k/N) = -X(k)\, \Lambda^{-1}(N)\, \lambda_p(N) \tag{A.34a}$$

Note that for $k \leq N$, exactly the same relation holds for $N - 1$ so

$$\hat{x}_h(k/N - 1) = -X(k)\, \Lambda^{-1}(N - 1)\, \lambda_p(N - 1) \tag{A.34b}$$

When (A.34a) is used in (A.15),

$$\hat{x}(k/N) = \hat{x}_p(k) - X(k)\, \Lambda^{-1}(N)\, \lambda_p(N) \tag{A.35}$$

But $\lambda_p(N)$ in the last equation is simply the particular solution to (A.13) starting with the initial conditions (A.14) and hence can be written

$$\begin{aligned} \lambda_p(N) &= \Phi_N^{-T}\left(I + S_N G_{N-1} Q_{N-1} G_{N-1}^T\right) \lambda_p(N - 1) \\ &\quad + \Phi_N^{-T} H_N R_N^{-1} \left\{ y_N - H_N\left[\Phi_{N-1} \hat{x}_p(N - 1) + u_{N-1}\right]\right\} . \end{aligned} \tag{A.36}$$

Next we use (A.24) to solve for $\hat{x}_p(N-1)$, substitute the result into (A.36), and collect terms [also using (A.13)] to show

$$\begin{aligned} \lambda_p(N) &= \Phi_N^{-T}\left[I + S_N G_{N-1} Q_{N-1} G_{N-1}^T - S_N \Phi_{N-1} X(N - 1)\, \Lambda^{-1}(N - 1)\right] \lambda_p(N - 1) \\ &\quad + \Phi_N^{-T} H_N R_N^{-1}\left[y_N - H_N \hat{x}(N/N - 1)\right] . \end{aligned} \tag{A.37}$$

When the Λ^{-1} term is factored out of the first bracket in (A.37) the remainder can be recognized as the homogeneous equation (A.13) for Λ. Hence

$$\lambda_p(N) = \Lambda(N)\Lambda^{-1}(N - 1)\lambda_p(N - 1) + \Phi_N^{-T}H_N^T R_N^{-1}\left[y_N - H_N\hat{x}(N/N - 1)\right] \quad (A.38)$$

(A.38) is now substituted in (A.35) to get the recursive relation

$$\hat{x}(k/N) = \hat{x}_p(k) - X(k)\Lambda^{-1}(N - 1)\lambda_p(N - 1)$$

$$- X(k)\Lambda^{-1}(N)\Phi_N^{-T}H_N^T R_N^{-1}\left[y_N - H_N\hat{x}(N/N - 1)\right] \quad (A.39)$$

$$= \hat{x}(k/N - 1) + W(k/N)\, H_N^T R_N^{-1}\left[y_N - H_N\hat{x}(N/N - 1)\right]$$

where we have defined

$$W(k/N) = -X(k)\Lambda^{-1}(N)\Phi_N^{-T} \quad . \quad (A.40)$$

We recall the definition (A.27) so

$$W(k/k) = -X(k)\Lambda^{-1}(k)\Phi_k^{-T}$$

$$= P_k \quad (A.41)$$

A recursive update equation is readily obtained for W by evaluating (A.40) in terms of $\Lambda(N-1)$ from (A.13):

$$W(k/N) = -X(k)\left[\left(I + S_N G_{N-1} Q_{N-1} G_{N-1}^T\right)\Lambda(N - 1) - S_N \Phi_{N-1} X(N - 1)\right]^{-1}$$

$$= -X(k)\Lambda^{-1}(N - 1)\Phi_{N-1}^{-T}\Phi_{N-1}^T\left[I + S_N\left(G_{N-1}^T Q_{N-1} G_{N-1} + \Phi_{N-1} P_{N-1} \Phi_{N-1}^T\right)\right]$$

$$= W(k/N - 1)\Phi_{N-1}^T(I + S_N P_{N/N-1})^{-1} \quad (A.42)$$

$$= W(k/N - 1)\Phi_{N-1}^T(I - S_N P_N)$$

Equations (A.39), (A.41) and (A.42) combined with the forward filtering equations (A.27), (A.28), (A.31) and (A.32) constitute a forward, fixed point smoothing algorithm.

Alternate Smoothing Equations

There are a variety of equations which can be used to obtain smoothed estimates satisfying (A.9) and (A.10) once the filtering solutions have been obtained. By storing filtered estimates only one $n \times 1$ matrix equation, rather than the two equations indicated need be solved backwards. The first step in this procedure is to use Eqs. (A.19) and (A.20) to show

$$\lambda_h(k/N) = \Lambda(k)X^{-1}(k/0) \tag{A.33}$$

To get an expression for the particular solution to λ Eq. (A.29) is solved, yielding

$$\lambda_p(k) = \Phi_k^{-T}P_k^{-1}\left[\hat{x}(k/k) - \hat{x}_p(k)\right] \tag{A.44}$$

Finally, Eqs. (A.43) and (A.44), combined with the defining relation (A.25) can be substituted into (A.15) to obtain

$$\begin{aligned}\lambda(k/N) &= \Phi_k^{-T}P_k^{-1}\left[-x_h(k/N) + \hat{x}(k/k) - \hat{x}_p(k/N)\right] \\ &= \Phi_k^{-T}P_k^{-1}\left[\hat{x}(k/k) - \hat{x}(k/N)\right]\end{aligned} \tag{A.45}$$

A simple backward smoothing equation is obtainable by solving (A.9) for $\hat{x}(k/N)$ and substituting (A.45) into the result. This gives

$$\hat{x}(k/N) = \Phi_k^{-1}\left\{\hat{x}(k + 1/N) - u_k + G_kQ_kG_k\left(\Phi_k^{-T}P_k^{-1}\right)\left[\hat{x}(k/k) - \hat{x}(k/N)\right]\right\} \quad . \tag{A.46}$$

This last equation is now rearranged and written as

$$\Phi_k^{-1}C_k^{-1}\hat{x}(k/N) = \Phi_k^{-1}C_k^{-1}\hat{x}(k/k) - \Phi_k^{-1}\left[\Phi_k\hat{x}(k/k) + u_k - \hat{x}(k + 1/N)\right] \quad (A.47)$$

where

$$\begin{aligned} C_k^{-1} &= \left(\Phi_k + G_kQ_kG_k\Phi_k^{-T}P_k^{-1}\right) \\ &= \left(\Phi_kP_k\Phi_k^T + G_kQ_kF_k^T\right)\Phi_k^{-T}P_k^{-1} \\ &= P_{k+1/k}\,\Phi_k^{-T}P_k^{-1} \quad . \end{aligned} \quad (A.48)$$

When (A.47) is solved we get the desired backward recursive relation

$$\begin{aligned} \hat{x}(k/N) &= \hat{x}(k/k) - C_k\left[\hat{x}(k + 1/k) - \hat{x}(k + 1/N)\right] \\ C_k &= P_k\Phi_k^TP_{k+1/k} \end{aligned} \quad (A.49)$$

Appendix B

The preceding appendix developed a number of equations related to generating least square state estimates for linear systems. This appendix briefly derives some of the statistical properties of those estimates utilized in Chapter 3.

The first such property is that the quantities $P_{k+1/k}$ and P_{k+1} actually represent the covariance of the estimate error $\tilde{x}_{k+1/k}$ and $\tilde{x}_{k+1}$ defined by

$$\tilde{x}_{k+1/k} = x_{k+1} - \hat{x}_{k+1/k} \tag{B.1a}$$

and

$$\tilde{x}_{k+1} = x_{k+1} - \hat{x}_{k+1} \tag{B.1b}$$

The true state is assumed to come from

$$x_{k+1} = \Phi_k x_k + G_k q_k + u_k \tag{B.2a}$$

with u_k known. Observations are generated according to

$$y_k = H_k x_k + \nu_k \tag{B.2b}$$

The three random variables in this problem are assumed to be the two noise sequences q_k and ν_k and the initial error $\tilde{x}_0 = x_0 - \bar{x}_0$. All these random variables are taken as being uncorrelated with zero mean. Their covariances are

$$\begin{aligned} E\left\{\tilde{x}_0 \tilde{x}_0^T\right\} &= P_0 \\ E\left\{\nu_k \nu_k^T\right\} &= R_k \\ E\left\{q_k q_k^T\right\} &= Q_k \end{aligned} \tag{B.3}$$

The proof that P is the covariance discussed above is based on an inductive argument. It is true for $k = 0$ by hypothesis. We now assume it is true for k and show that it is true for $k + 1$. Differencing (A.32) and (B.2a) gives

$$\tilde{x}_{k+1} = \Phi_k \tilde{x}_k + G_k q_k \tag{B.4}$$

so that

$$\begin{aligned} E\left\{\tilde{x}_{k+1/k} \tilde{x}^T_{k+1/k}\right\} &= \Phi\, E\left\{\tilde{x}_k \tilde{x}^T_k\right\} \Phi^T_k + G_k Q_k G^T_k \\ &= \Phi_k P_k \Phi^T_k + G_k Q_k G^T_k \\ &= P_{k+1/k} \end{aligned} \tag{B.5}$$

Similarly, with the aid of (B.2b) and (A.32) equations (A.32), (A.31) and (B.2a) are differenced to get

$$\tilde{x}_{k+1/k+1} = (I - P_{k+1} S_{k+1}) \tilde{x}_{k+1/k} - P_{k+1} H_{k+1} R^{-1}_{k+1} \nu_{k+1} \tag{B.6}$$

The covariance of this quantity is readily seen to be

$$\begin{aligned} E\left\{\tilde{x}_{k+1/k+1} \tilde{x}^T_{k+1/k+1}\right\} &= (I - P_{k+1} S_{k+1})\, E\left\{\tilde{x}_{k+1/k} \tilde{x}^T_{k+1/k}\right\} (I - P_{k+1} S_{k+1})^T \\ &\quad + P_{k+1} S_{k+1} P_{k+1} \\ &= (I - P_{k+1} S_{k+1})\, P_{k+1/k} (I - P_{k+1} S_{k+1})^T \\ &\quad + P_{k+1} S_{k+1} P_{k+1} \end{aligned} \tag{B.7}$$

From Eq. (A.28),

$$P_{k+1/k} = (I - P_{k+1} S_{k+1})^{-1} P_{k+1}$$

which, when substituted into (B.7) gives

$$E\left\{\tilde{x}_{k+1/k+1}\tilde{x}^T_{k+1/k+1}\right\} = P_{k+1} \tag{B.8}$$

which was to be shown.

Next some useful outer product orthogonality relations will be derived. The first of these is that

$$E\left\{\hat{x}_i\tilde{x}_i^T\right\} = 0$$

$$i = 0, \ldots N \tag{B.9}$$

This relation is also proved by induction. In fact, by hypothesis

$$E\left\{\hat{x}_0(x_0-\hat{x}_0)^T\right\} = \bar{x}_0 E\left\{x_0^T\right\} - \bar{x}_0\bar{x}_0^T = 0 \tag{B.10}$$

We now assume (B.9) is true for i and show this implies its truth for $i = 1$. In fact

$$E\left\{\hat{x}_{i+1}\tilde{x}^T_{i+1}\right\} = E\left\{\left[\Phi_i\hat{x}_i + P_{i+1}S_{i+1}(\Phi_i\tilde{x}_i + G_iq_i) + P_{i+1}H^T_{i+1}R^{-1}_{i+1}\nu_{i+1}\right]\right.$$

$$\left.\cdot\left[\tilde{\Phi}_i\tilde{x}_i + (I - P_{i+1}S_{i+1})G_iq_i - P_{i+1}H^T_{i+1}R^{-1}_{i+1}\nu_{i+1}\right]^T\right\} \tag{B.11}$$

When (B.11) is expanded and only the non zero expectations retained,

$$\begin{aligned}
E\left\{\hat{x}_{i+1}\tilde{x}^T_{i+1}\right\} &= P_{i+1}S_{i+1}\Phi_i E\left\{\tilde{x}_i\tilde{x}_i^T\right\}\tilde{\Phi}_i^T \\
&\quad + P_{i+1}S_{i+1}G_i\ E\left\{q_iq_i^T\right\}\ G_i^T(I - P_{i+1}S_{i+1})^T \\
&\quad - P_{i+1}H^T_{i+1}R^{-1}_{i+1}\ E\left\{\nu_{i+1}\nu^T_{i+1}\right\}\ R^{-1}_{i+1}H_{i+1}P_{i+1} \\
&= P_{i+1}S_{i+1}\left[\Phi_iP_i\Phi_i^T + G_iQ_iG_i^T\right](I - P_{i+1}S_{i+1})^T - P_{i+1}S_{i+1}P_{i+1}
\end{aligned} \tag{B.12}$$

The term is brackets is identified as $P_{i+1/i}$ and

$$P_{i+1/i}(I - S_{i+1}P_{i+1}) = P_{i+1}$$

so

$$E\left\{\hat{x}_{i+1}\tilde{x}^T_{i+1}\right\} = P_{i+1}S_{i+1}P_{i+1} - P_{i+1}S_{i+1}P_{i+1} = 0$$

which completes the proof of (B.9).

Having derived (B.9) it is easy to show that

$$E\left\{\hat{x}_j\tilde{x}^T_i\right\} = 0 \qquad i \geq j \tag{B.13}$$

This is deduced by simply writing

$$\tilde{x}_i = \tilde{\Phi}(i+j)\tilde{x}_j + \sum_{k=j}^{i-1} \tilde{\Phi}_{i-1/k} \left[(I - P_{k+1}S_{k+1})G_k q_k - P_{k+1}H^T_{k+1}R^{-1}_{k+1}\nu_{k+1}\right] \tag{B.14}$$

By hypothesis, $\tilde{x}_j$ is normal to $\tilde{x}_j$, and independent of the g_k and ν_{k+1} so when (B.14) is substituted into (B.13) the result is proven.

It is slightly more difficult to show that

$$E\left\{\hat{x}_j\tilde{x}^T_i\right\} = 0 \qquad i \leq j \tag{B.15}$$

This requires an induction argument. By hypothesis, it is true for $j = i$. We assume it is true for some $k > i$ and show this implies it is true for $k + 1 > i$. First, the estimate $\hat{x}_{k+1}$ is written in closed form as

$$\begin{aligned}\hat{x}_{k+1} &= \Phi_{(k+1/i)}\hat{x}_i \\ &+ \sum_{\ell=i}^{k} \Phi_{(k/\ell)} \left[P_{\ell+1}H^T_{\ell+1}R^{-1}_{\ell+1}(H_{\ell+1}\Phi_\ell\tilde{x}_\ell + H_{\ell+1}G_\ell q_\ell + \nu_{\ell+1}\right]\end{aligned} \tag{B.16}$$

When (B.16) is inserted into the left side of (B.15) we see that the only term which can contribute to the expectation is the correlation between $\tilde{x}_\ell$ and $\tilde{x}_i^T$. But by hypothesis these expectations are zero so the result is proven.

It is now easy to prove the two relations used in Chapter 3:

$$E\left\{\hat{x}_{\ell/N}\tilde{x}^T_{\ell/N}\right\} = 0 \qquad \ell \leq N \tag{B.17a}$$

$$E\left\{\hat{x}_{\ell/N}\tilde{x}_{\ell}\right\} = 0 \qquad \ell \leq N \tag{B.17b}$$

The smoothed quantities are just written as sums of the filtered quantities using (A.49), and relations (B.13) and (B.15) used to show all the terms give zero expectation.

Chapter 3 also uses the relation

$$E\left\{\hat{x}_{\ell/N}\nu^T_{\ell}\right\} = P_{\ell/N}H^T_{\ell} \ . \tag{B.18}$$

One way to get this expectation is to use the forward update equation (A.39). Before doing this, however, it is necessary to demonstrate that the smoothed covariance obeys the forward update equation.

$$P_{\ell/j} = P_{\ell/j-1} - W(\ell/j)\left[S_j + S_jP_{j/j-1}S_j\right]W^T(\ell/j)$$

$$N \geq j \geq \ell + 1 \tag{B.19a}$$

with initial conditions

$$W(\ell/\ell) = P_\ell \triangleq P_{\ell/\ell} \tag{B.19b}$$

The update for W is given by (A.42). In error equation form (A.40) becomes

$$\tilde{x}_{\ell/j} = \tilde{x}_{\ell/j-1} - W(\ell/j)H_j^T R_j^{-1}(H_j \tilde{x}_{j/j-1} + \nu_j)$$

$$j = \ell + 1,\ \ell + 2,\ \ldots\ N \qquad \text{(A.20)}$$

When each side of (B.20) is multiplied on the right by its transpose and the expectation taken, we get

$$P_{\ell/j} = P_{\ell/j-1} + W(\ell/j)\ S_j P_{j/j-1} S_j + W(\ell/j) S_j W^T(\ell/j)$$

$$- W(\ell/j) S_j\ E\left\{\tilde{x}_{j/j-1} \tilde{x}^T_{\ell/j-1}\right\} - E\left\{\tilde{x}_{\ell/j-1} \tilde{x}^T_{j/j-1}\right\} S_j W^T(\ell/j)$$

$$\text{(B.21)}$$

Using (B.21), we will now show that (B.19a) is true for $j = \ell + 1$ and that its truth for any $j-1$ implies its truth for j. For $j = \ell + 1$, (B.21) becomes

$$P_{\ell/\ell+1} = P_{\ell/\ell} + W(\ell/\ell+1)(S_{\ell+1} P_{\ell+1/\ell} S_{\ell+1} + S_{\ell+1}) W^T(\ell/\ell+1)$$

$$- W(\ell/\ell+1)\ S_{\ell+1} \Phi_\ell P_\ell - P_\ell \Phi_\ell^T S_{\ell+1} W^T(\ell/\ell+1) \qquad \text{(B.22)}$$

The last two terms in (B.22) can be combined with the preceding terms by using (A.42). For example

$$W(\ell/\ell+1) S_{\ell+1} \Phi_\ell P_\ell = W(\ell/\ell+1) S_{\ell+1} (I + P_{\ell+1/\ell} S_{\ell+1})(I + P_{\ell+1/\ell} S_{\ell+1})^{-1} \Phi_\ell$$

$$= W(\ell/\ell+1)(S_{\ell+1} + S_{\ell+1} P_{\ell+1/\ell} S_{\ell+1}) W^T(\ell/\ell+1) \qquad \text{(B.23)}$$

Thus (B.22) can be rewritten

$$P_{\ell/\ell+1} = P_\ell - W(\ell/\ell+1)(S_{\ell+1} P_{\ell+1/\ell} S_{\ell+1} + S_{\ell+1}) W^T(\ell/\ell+1) \qquad \text{(B.24)}$$

which proves that (B.19a) is true for $j = \ell + 1$. We now assume that (B.19a) is true for any $j-1$ and prove its truth for j. We do this by using (B.21) again. First, however, we induce that

$$E\left\{\tilde{x}_{\ell/j-1}\tilde{x}^T_{j/j-1}\right\} = W(\ell/j)(I + S_j P_{j/j-1} \tag{B.25}$$

relying on the fact that (B.19b) is in fact true up to some $j-1$. The truth of (B.25) for $j = \ell + 1$ was essentially demonstrated during the proof of (B.23). We assume (B.25) is true for $j-1$ so

$$E\left\{\tilde{x}_{\ell/j-2}\tilde{x}^T_{j-1/j-2}\right\} = W(\ell/j-1)\left(I + S_{j-1}P_{j-1/j-2}\right) \tag{B.26}$$

Using (B.20) and the update of $\tilde{x}_{j-1/j-2}$ we find

$$E\left\{\tilde{x}_{\ell/j-1}\tilde{x}^T_{j/j-1}\right\} = E\left\{\left[\tilde{x}_{\ell/j-2} - W(\ell/j-1)H^T_{j-1}R^{-1}_{j-1}\left(H_{j-1}\tilde{x}_{j-1/j-2} + \nu_{j-1}\right)\right]\right.$$

$$\left.\cdot\left[G_{j-1}q_{j-1} + \Phi_{j-1}(I - P_{j-1}S_{j-1})\tilde{x}_{j-1/j-2} - \Phi_{j-1}P_{j-1}H^T_{j-1}R_{j-1}\right]\right\}$$

$$= E\left\{\tilde{x}_{\ell/j-2}\tilde{x}_{j-1/j-2}\right\}(I - S_{j-1}P_{j-1})\Phi^T_{j-1}$$

$$= W(\ell/j-1)(I + S_{j-1}P_{j-1/j-2})(I - S_{j-1}P_{j-1})\Phi^T_{j-1}$$

$$= W(\ell/j-1)\Phi^T_{j-1}(I + S_jP_{j/j-1})^{-1}(I + S_jP_{j/j-1})$$

$$= W(\ell/j)(I + S_jP_{j/j-1}) \tag{B.27}$$

so that (B.25) is proved. The proof of (B.19a) is concluded by substituting (B.25) into (B.21) to complete the induction.

With (B.19) proven it is now simple to prove (B.18). For $N = \ell$ the relation is clearly true, as can be shown by using the filtering relation for $\hat{x}_{\ell/\ell}$ as a function of $\hat{x}_{\ell/\ell+1}$. We now assume it is true for any $N-1$ and show this implies it is true for N. In fact, using (A.39),

$$E\left\{\hat{x}_{\ell/N}\nu_\ell^T\right\} = E\left\{\left[\hat{x}_{\ell/N-1} + W(\ell/N)H_N^T R_N^{-1}(H_N x_N + \nu_N - H_N\hat{x}_{N/N-1})\right]\nu_\ell^T\right\}$$

$$= P_{\ell/N}H_\ell^T - W(\ell/N)S_N\ E\left\{\hat{x}_{N/N-1}\nu_\ell^T\right\}\ . \tag{B.28}$$

Using the filtering relations starting with

$$\hat{x}_{\ell+1/\ell} = \Phi_\ell\left[\hat{x}_{\ell/\ell+1} + P_\ell H_\ell^T R_\ell^{-1}(H_\ell x_\ell - H\hat{x}_{\ell/\ell} + \nu_\ell)\right]$$

we can see that

$$E\ \hat{x}_{\ell+1/\ell}\nu_\ell^T = \Phi_\ell P_\ell H_\ell^T$$

By induction, using these relations, it can be shown that

$$E\left\{\hat{x}_{N/N-1}\nu_\ell^T\right\} = \Phi_{N-1}W^T(\ell/N-1)H_\ell^T$$

Hence, (B.28) becomes

$$E\left\{\hat{x}_{\ell/N}\nu_\ell^T\right\} = (P_{\ell/N-1} - W(\ell/N)S_N\Phi_{N-1}W^T(\ell/N-1)H_\ell^T$$

$$= \left[P_{\ell/N-1} - W(\ell/N)(S_N + S_N P_{N/N-1}S_N)W(\ell/N)\right]H_\ell^T$$

$$= P_{\ell/N}H_\ell^T\ .$$

BIBLIOGRAPHY

1. T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.

2. Michael Athans, Richard P. Wishner, and Anthony Bertolini, "Suboptimal state estimation for continuous-time non-linear systems from discrete noisy measurements," Trans. Joint Automatic Control Conference, June, 1968, pp 364-382.

3. M. S. Bartlett, An Introduction to Stochastic Processes, The University Press, Cambridge, England, 1960.

4. R. W. Bass, V.D. Norum, and L. Schwartz, "Optimal multichannel non-linear filtering," Journal of Math. Analysis and Applications, vol. 16, 1966, pp 152-164.

5. J. V. Breakwell, "Slightly non-linear estimation with noisy data," Proc. of AIAA Guidance Control and Flight Mechanics Conference, Princeton, N. J., August, 1969.

6. Arthur E. Bryson and Malcolm Frazier, "Smoothing for linear and non-linear dynamic systems," Proc. Optimum Sys. Synthesis Conference, U.S. Air Force Tech. Rept., ASD-TDR-63-119, Feb., 1963.

7. Arthur E. Bryson, Jr. andYu-Chi Ho, Applied Optimal Control, Blaisdell Publishing, Waltham, Mass., 1969.

8. Bjorn Conrad, "The effect of small non-linearities on several common state estimation schemes," Proc. of Symposium on Non-linear Estimation Theory and its Applications, Sept., 1970, pp. 263-265.

9. Andrew H. Jazwinski, "Nonlinear filtering with discrete observations," AIAA paper 66-38, presented at the AIAA 3rd Aerospace Sciences Meeting, New York, Jan., 1966.

10. R. E. Kalman, "A new approach to filtering and prediction problems," Trans. ASME J. Basic Engrg., ser. D, vol. 82, 1960, p. 35.

11. R. E. Kalman and R. S. Busy, "New Results in linear filtering and prediction theory," Trans. ASME J. Basic Engrg., ser. D., vol. 83, March, 1961, pp. 95-108.

12. J. S. Meditch, "Optimal Fixed-Point Continuous Linear Smoothing," Proc. of JACC, University of Pennsylvania, June, 1967.

13. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill Co., New York, 1965.

14. H. E. Rauch, F. Tung, and C. T. Strubel, "Maximum likelihood estimates of linear dynamic systems," AIAA Journal, vol. 3, 1965, p 1445.

15. L. Schwartz, "Approximate Continuous Non-linear Minimal-Variance Filtering," UCLA Report 67-17, April, 1967.

16. C. K. Lee, Optimal Estimation, Identification, and Control, The MIT Press, Cambridge, Mass., 1964.

17. P. Whittle, Prediction and Regulation by Linear Least Square Methods, D. Van Nostrand Co., Princeton, N.J., 1963.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

Dept. of Aero and Astro
Stanford Univ., Stanford, Calif. 94305

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

STATE ESTIMATION WITH SMALL NON-LINEARITIES

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Björn Conrad

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1971	94	17

8a. CONTRACT OR GRANT NO.

F33615-C-67-1245 &
F33615-C-70-1637

NASA NGR-05-020-019

b. PROJECT NO.

c. Task 510215

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

SUDAAR No. 421

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

AFAL-TR-71- 222

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

AF Avionics Lab(AFSC)
Wright-Patterson AFB
Ohio 45433

Code SG-NASA HDQS
Washington 25,D.C.

13. ABSTRACT

A variety of techniques are available for estimating the states of non-linear dynamic systems from noisy data. These procedures are generally equivalent when applied to linear systems. This report investigates the difference between several of these procedures in the presence of small dynamic and observational non-linearities.

Four discrete estimation algorithms are analyzed. The first is a strictly least square estimator, while the other three are recursive algorithms similar to the Kalman filter used for estimating the states of linear systems. The product of this research is a group of analytic expressions for the mean and covariance of the error in each of those estimators so that they may be compared without lengthy Monte-Carlo simulations.

The covariance expressions show that, to first order, all the estimators have the same covariance. Expressions for the means, however, show that each estimator has a different bias. Several examples are carried out demonstrating that the relative magnitudes of the bias errors in the various estimators can be a strong function of such parameters as initial covariances and number of data points being considered. In fact, under some circumstances it appears that more complicated (seemingly superior) algorithms can have larger biases than smaller ones.

DD FORM 1 NOV 65 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
NOISY DATA						
LINEAR SYSTEMS						
DISCRETE ESTIMATION ALGORITHMS						
LEAST SQUARE ESTIMATOR						
RECURSIVE ALGORITHMS						
KALMAN FILTER						
ESTIMATOR						

UNCLASSIFIED

Security Classification